分类号____________　　　　　　　　　　　　　　　　密级 ________
U D C ____________　　　　　　　　　　　　　　　　编号 10741

蘭州财經大學
LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕 士 学 位 论 文

论文题目 基于 Guass-Seidel 型迭代算法的分层模型统计推断

研 究 生 姓 名： 周梦雨

指导教师姓名、职称： 田茂再 教授

学 科、专 业 名 称： 统计学 数理统计学

研 究 方 向： 复杂数据分析

提 交 日 期： 2021 年 6 月 6 日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：周梦雨　　签字日期：2021年6月6日

导师签名：田兆彝　　签字日期：2021年6月6日

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，同意（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名：周梦雨　　签字日期：2021年6月6日

导师签名：田兆彝　　签字日期：2021年6月6日

Statistical Inference of Hierarchical Model Based on Guass -Seidel Type Iterative Algorithm

Candidate :Zhou Mengyu

Supervisor:Tian Maozai

摘 要

缺失数据现象在观测数据中是十分常见的问题，而EM算法是针对缺失数据问题中求参数估计的常用方法，是将不完全数据转化为完全数据问题来处理。针对分层线性回归模型中参数估计问题，常常以两层数据模型为例，使用EM算法迭代求解。本文针对缺失数据的分层线性回归模型提出了Guass-Seidel型迭代方法，其主要思想是，迭代产生当前最新值，并运用最新值计算得到下一步最新参数估计值，对分层线性模型中固定效应与随机效应部分的参数估计进行推导，依据判断准则决定迭代过程的收敛，直至迭代过程结束。本文主要贡献之一在于利用Guass-Seidel型迭代算法提高参数迭代收敛速度。

对于处理统计推断中假设检验问题，统计推断是通过样本数据的观测信息来推断总体的主要方法，对于具有嵌套结构的分层线性回归模型的系数向量诊断方法，对分层线性回归模型的第一层模型系数诊断主要利用传统的线性嵌套回归模型F检验进行统计推断。本文在对分层线性回归模型的第二层系数进行统计诊断中，将具有嵌套结构的多元线性回归模型推广到具有嵌套结构的分层线性回归模型中，主要构建分层线性回归模型似然函数比值来构造检验统计量，来判断第二层系数向量的拟合问题。

最后通过随机模拟的数值结果表明，与文献中的分层线性模型的 EM 算法相比，Guass-Seidel 型迭代方法具有更有效的收敛速度。通过高校数学成绩数据来说明具有嵌套结构的分层线性模型的统计推断方法的有效性和实用性。

关键词：分层线性模型 EM 算法 Guass-Seidel 型迭代 嵌套模型 统计推断

Abstract

Missing data phenomenon is a very common problem in observation data, and EM algorithm is a common method for parameter estimation in missing data problem, which transforms incomplete data into complete data problem to deal with. For parameter estimation problem in Hierarchical Linear Regression Model, we often take two-level data model as an example, and use EM algorithm to solve iteratively. The main idea of this thesis is to generate the current latest value by iteration, and use the latest value to calculate the next step of the latest parameter estimation. The parameter estimation of fixed effect and random effect in Hierarchical Linear Model is derived, and the convergence of the iterative process is determined according to the judgment criteria until the end of the iterative process. The main contribution of this thesis is as follows One is to use the Guass-Seidel type iterative algorithm to improve the convergence speed of parameter iteration.

To deal with the problem of hypothesis testing in statistical inference, statistical inference is the main method to infer the population through the observation information of sample data. For the coefficient vector diagnosis method of Hierarchical Linear Regression Model with nested structure, the traditional F-test of Linear Nested Regression Model is mainly used to diagnose the coefficient of the first level model of Hierarchical Linear Regression Model. In this thesis, in the statistical diagnosis of the second layer coefficient of Hierarchical Linear Regression Model, we extend the multiple Linear Regression Model with nested structure to the Hierarchical Linear Regression Model with nested structure. We mainly construct the likelihood function ratio of the Hierarchical Linear Regression Model to construct the test statistics to judge the fitting problem of the second layer coefficient vector.

Finally, the numerical results of stochastic simulation show that the Guass-Seidel type iterative method has more effective convergence rate than the EM algorithm of Hierarchical Linear Model in literature. The validity and practicability of the statistical inference method of Hierarchical Linear Model with nested structure are illustrated by the data of college mathematics scores.

Keywords: Hierarchical Linear Model; EM algorithm; Guass-Seidel iteration; Nested model; Statistical inference

目　录

1 引 言

1.1 研究背景及意义

统计推断是利用样本提供的信息推断总体分布或总体的某些数字特征，从而认识总体特征的方法。统计推断分为两大类：一类是参数估计，另一类是假设检验。对于参数估计，主要分为点估计和区间估计，点估计常见方法有矩估计法和最大似然估计法。这些传统的参数估计方法一般依赖于完全数据。然而，在统计数据中，存在大量的缺失数据，造成数据缺失的原因有很多，如获取信息代价太大，人为失误，机器故障等。对存在缺失数据进行参数估计，一般方法是通过某种方式对缺失数据进行近似处理，对处理后的数据再利用传统的参数估计方法进行参数估计。对于缺失数据常见的处理方法有，直接删除含有缺失数据的记录或者对缺失值进行插补，或者是不处理，但一般方法是将不完全数据转化为完全数据来处理，这些方法在不同的应用中各有优势和缺陷。对于缺失情况下的参数估计问题和传统极大似然估计问题已不能直接解决，EM 算法的提出，对这一问题进行了弥补。

EM 算法是针对不完全数据进行参数估计求解的迭代算法，是在观测数据的基础上引进不可观测的“潜在数据”，从而将复杂的不完全数据问题转化为完全数据问题, 并完成一系列的简单的计算，如极大值或模拟。“潜在数据”可理解为观测数据所含有的缺失值所对应的变量，也可以理解为观测数据本身不存在缺失，为方便求解参数值而引进的隐变量。在求解参数估计时 EM 算法十分方便，针对分层线性回归模型的参数求解过程，利用 EM 算法求解过程中存在 E 步显式难实现的问题或者是收敛速度缓慢的问题。针对这些问题，提出相应的算法提高收敛速度的文献较少，为了方便实际问题中的应用，需要考虑收敛速度和收敛准则的问题，本文针对多层线性回归模型的参数估计问题给出了具体的方法。同时针对嵌套结构的多元线性模型的统计推断的文献较多，例如给出具有嵌套结构的多元线性回归模型形式以及针对具有嵌套结构多元线性回归模型给出的检验方法，考虑F检验，通过残差平方和构造检验统计量，去判断具有嵌套结构的多元线性回归模型拟合优度问题，但是针对具有嵌套结构的分层线性模型的统计推断的参考文献较少，对于具有嵌套结构的分层线性模型的检验方法文献较少。

在本文中，由于分层线性模型的协方差变量允许有异方差的情况，本文主要考虑同方差情况下。通常情况在处理观测数据过程中发现数据往往具有层次结构，这时使用传统的单水平分析模型不足以解释问题，就要考虑使用多水平模型解决问题。对于多层模型是建立在更合理的假设情况下的，考虑了来自不同层次的随机误差和变量信息，能够提供更加准确的标准误估计和更有效的区间估计和假设检验，同时多层线性模型可以计算任何水平上测量的协方差。多层次分析不仅可以用于分析观测变量之间的因果关系，而且可以作为协变量方差结构模型的扩展。对于分层线性模型中未知参数的估计是非常重要的，针对存在缺失数据分层线性模型的参数估计，以往的文献提出使用 EM 算法迭代来进行参数估计，但是 EM 算法往往存在收敛速度缓慢的特征，通常都是通过对 EM 算法的 E 步进行创新，来解决 E 步难求解的问题，对提高收敛速度做出贡献的文献较少。使用 Guass-Seidel 型迭代算法提高并行算法的效率在求解线性方程组过程是十分常见的，其主要思想是其迭代过程中利用迭代产生的最新值，将其思想引用到分层线性模型 EM 算法的迭代过程中，提出分层线性模型 Guass-Seidel 型迭代算法，并给出了具体的迭代过程，讨论分层线性模型 Guass-Seidel 型迭代算法在参数估计的收敛速度上与分层线性模型 EM 算法相比是否具有更好的收敛性，例如通过对比迭代次数，观测收敛参数值等方法。

同时似然比检验方法是在复合假设的检验问题中构造检验的常用方法，在具有嵌套结构的分层线性回归模型中进行统计推断，其中对第一层模型系数诊断即是嵌套结构的传统多元线性模型的检验问题，可以使用多种方法进行，例如：F检验，t检验，或者判定系数来解释拟合优度问题。对于具有嵌套结构的多层线性回归模型的统计推断问题，即形式上为第二层模型上引入系数向量，使在多层线性回归模型上具有嵌套结构。基于同方差的情况下，对第二层分层线性模型的系数进行统计推断，主要利用构造似然比的方法, 判断对引入第二层变量时是否保留的问题给出理论方法，将复杂的判定第二层模型系数向量问题简单化。针对分层线性回归模型的统计推断问题不仅可以求解参数估计问题，还给出了分层模型中假设检验的一般方法。本文研究的意义如下：

在许多实际应用数据中，往往都存在分层数据结构，但在过去的研究往往未能在数据分析中充分解决它们，在很大程度上，这种忽视反映了传统统计技术对

嵌套结构线性模型估计的局限性。使用多层次模型处理数据可以观测到不同水平下的具体信息，可以很容易地提出关于在每个层次和跨层次发生的假设，也可以评估每个层次上的变化量。从实质性的角度来看，多层次模型更符合许多行为和社会研究中所研究的基本现象。EM 算法或者迭代广义最小二乘法是常用的分层线性回归模型的参数求解方法，但这些算法的性能包括它的收敛速度和收敛的可靠性是不同的，然而 EM 算法在分层线性模型参数估计过程中往往存在收敛速度缓慢的问题。Guass-Seidel 型迭代算法在求解线性方程组理论中是相当成熟的，其迭代产生最新值的思想可以提高并行算法的效率，在实际应用中可以减少处理机间的通信次数，所以基于在分层线性模型 EM 算法的基础上引入 Guass-Seidel 型迭代算法是具有实际意义和现实意义的。同时具有嵌套结构的线性模型理论是非常丰富的，对于处理具有嵌套结构的线性模型检验统计量的应用是十分广泛的，同时似然比在统计学应用中是同时反映特异度和灵敏度的复合指标，构造似然比统计量在检验模型是否符合数据分析结果在实际应用中是非常成熟的。但是针对具有嵌套结构的分层线性模型的理论依据和判断准则相对较少，利用构造似然比统计量的方法来作为判断具有嵌套结构的分层线性模型的理论依据是具有一定实践意义的。

1.2 国内外研究现状

在实际观测数据中，数据往往具有分层结构，对于分层线性回归模型的描述，不同的文献中名称是有差异的。在社会学研究中，通常被称为多层线性模型(Goldstein[1],1995)，在生物计量学应用中，称为随机效应模型(Laird[2],1982)，在统计学文献中，一般称为协方差成分模型(Dempster[2],1981)。同时观测数据普遍存在缺失数据，Little 和 Rubin[4](1987)将缺失数据分为三种类型，本文针对随机缺失类型问题进行研究，在观测数据存在缺失的情况下加入随机缺失数据，形成完全数据，然后对完全数据进行参数估计的过程。Dempster[5](1977)介绍了在观测数据存在缺失数据时，通过极大似然估计算法进行迭代计算，其算法是由 E 步（期望步）和 M 步（最大值）两步组成，并给出了具体的迭代过程和大量的例子，其理论具有简单性和普遍性。Wu[6](1983)介绍了 EM 算法两个方面收敛性的问题。但是 EM 算法也存在一些不足的问题，例如缺失数据较高时，那么算法

的收敛速度就会比较慢，或者 EM 算法中 M 步比较难以实现，即对数似然函数难以给出准确公式，进行参数估计是比较困难的，或者算法中的 E 步积分难以实现。其中关于 EM 算法，Stephen 和 Raudenbush[7](1997)针对分层线性模型的特殊例子线性混合效应模型，解决待估参数的估计问题，用完全数据充分统计量的条件期望来估计完全数据充分统计量，来代替 E 步难表达的问题，并给出分层线性模型 EM 算法的一般迭代过程。Meng 和 Rubin[8](1993)基于 EM 算法收敛性稳定的基础上，提出了 ECM 算法来解决 EM 算法中 M 步难求解的问题。关于 EM 算法的应用，Louis[9] (1982)提出了加速 EM 算法收敛的方法，这些文献体现了 EM 算法在分层线性模型参数估计的可行性，同时保证了算法在分层线性模型上的收敛性。Vermunt[10](2010)实现了 EM 算法在两层以上嵌套数据集的分层非线性模型的极大似然估计；Walker[11](1996)在非线性随机效应模型中使用了蒙特卡洛模拟代替 E 步的方法，提高了 EM 算法的适用范围。基于嵌套结构的文献，针对统计推断过程，Chown 和 Müller [12](2019)介绍了一种多协变量非参数回归模型的异方差检验方法，基于残差的经验分布函数利用局部多项式平滑构造残差，设计检测函数来验证异方差。针对具有嵌套结构的模型检验问题，Alfandari[13]等(2021)研究了利用嵌套的 *Logit* 选择模型对顾客偏好进行建模，从而找到期望收益最大化的最优产品分类的问题。Li 和 Lee[14](2019)通过似然函数最大化拟合半参数零膨胀负二项分布回归模型，并通过似然比评价一个假设参数泛函形式的连续协变量效应的充分性，利用数据证明其方法的有效性。

国内针对 EM 算法，田茂再[15](2015)介绍了近年来国际上关于分层分位回归建模理论与方法的新成果，其中包含 EM 算法的基本内容，同时给出了协方差已知和未知情况下的 EM 计算方法，并在分层线性分位回归模型情况下，介绍了 Guass-Seidel 型迭代算法的过程，本文的思想是基于此理论上进行计算和推导的。Guass-Seidel迭代的主要思想是在迭代算法过程中充分利用最新值的方法，使迭代算法产生得到的参数更具有价值性。尚月强[16](2006)在求解线性方程组中利用 Guass-Seidel并行迭代算法，给出了在求解线性方程组过程中利用Guass-Seidel并行迭代算法的具体算法显式，提高了并行算法的效率。本文是基于这种迭代算法的思想上构建的，运用到分层线性回归模型中，提高EM算法的收敛速度。乔舰和景平[17](2012)给出了分层线性模型中固定效应部分和随机效应部分的具体参

数估计推导。卢玉桂[18](2013)针对EM算法中E步积分的式子表达难以实现，用MCEM算法来完成分层线性模型的参数估计，主要是通过蒙特卡洛模拟方法随机抽取随机数来解决E步积分难计算的问题，然后得到缺失变量的初值，带入M步完成迭代过程，直至参数收敛。贺建风等[19](2019)利用小域样本特征，将广义线性模型引入到分层贝叶斯方法中，解决调查中多层次推断问题。赵必华等[20](2019)和王芳[21](2018)针对多层次线性模型给出了实际的应用。

考虑分层线性回归模型的回归系数诊断问题，对分层线性回归模型第一层回归系数进行回归诊断，即对应的为多元线性回归系数诊断问题。在多元线性回归模型中，吴喜之[22](2016)用线性模型来近似因变量与自变量的线性关系，并给出线性模型中参数估计的方法和关于多自变量系数复合检验过程。谢宇[23](2013)介绍了多元线性模型中嵌套模型的具体含义，并给出了嵌套模型多个回归系数的联合检验常用的F统计量，同时给出二分因变量嵌套模型的模型评价方法，但并没有给出分层线性嵌套模型的检验过程。田茂再[24](2006)提出了基于 Guass-Seidel 迭代的条件分位分层线性回归模型的算法，解决了分层模型不能全面刻画高维情况下响应变量的条件分布问题。吴密霞[25](2013)介绍了模型参数的似然比检验，给出了具体的假设过程。针对具有嵌套结构的分层线性回归模型的系数诊断则有少量的参考文献，刘红云[26](2005)在追踪数据分析中介绍，对于多层线性模型，可以利用 *Wald* 检验来对固定部分参数进行显著性检验，同时利用整体拟合差异统计量来检验嵌套模型。

对于回归模型的统计推断推广，马海强[27](2008)给出变系数模型的统计诊断方法；戴林送[28](2013)研究了广义泊松回归模型的统计诊断方法；梁晋雯[29](2020)基于数据删失模型和均值漂移模型构建统计量进行异常点的诊断，研究体积抽样受异常点的影响。

1.3 文献评述

国外现有的文献研究集中在以下几方面：

（1）对 EM 算法的实际应用较多，提高迭代算法效率的文献研究较少。

（2）对于具有嵌套结构的分层线性模型提出了参数估计方法，但是对于嵌套结构的分层线性模型检验的文献却较少。

国内的相关文献可以发现：

（1）国内关于分层线性模型的参数估计理论较多，但是对分层线性模型的统计推断较少。

（2）Guass-Seidel 型迭代算法在求解线性代数中运用较多，但对于统计模型中参数估计的应用较少。

（3）对于分层线性模型第一层系数的统计推断即多元线性回归模型统计推断有成熟的理论依据，但是对于分层线性回归模型的第二层系数推断的研究较少。

1.4 文章结构

本文的文章结构内容可以分为以下的部分：

第一部分主要包括在随机缺失数据情况下提出的 EM 算法的研究发展现状以及 Guass-Seidel 型迭代算法的基本思想以及本文所出现的创新之处，并且介绍了本文的研究背景和国内外相关研究的文献综述，同时还给出了基本的文章框架和内容框架；

第二部分是本文对基础理论的梳理与总结，主要提出 EM 算法在随机缺失数据情况下的具体迭代过程以及 Guass-Seidel 型算法在求解线性方程组中的理论依据，并简单介绍了以两层分层线性模型为例的迭代算法的收敛准则，为文章后续的理论研究做了铺垫；

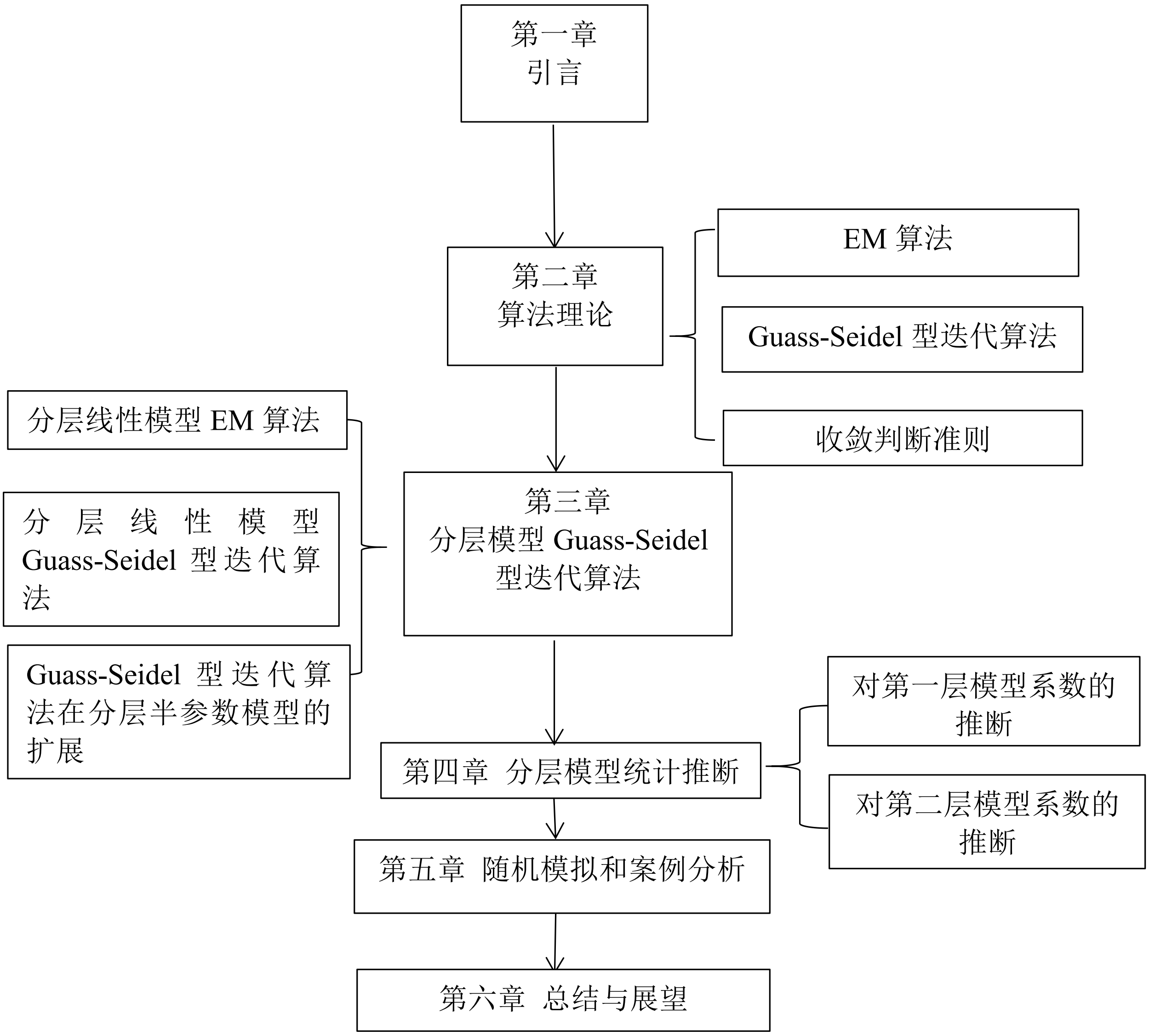
第三部分是本文的创新点之一，主要进行分层线性回归模型的 EM 算法的理论推导，并在 EM 算法的基础上，提出基于分层线性回归模型 Guass-Seidel 型迭代算法，并将 Guass-Seidel 型迭代算法扩展到分层半参数模型中，体现 Guass-Seidel 型迭代算法的一般性；

第四部分是本文的另一创新点，在分层线性嵌套模型的第一层系数进行统计推断即是对多元线性回归的统计推断的基础上，对第二层系数变量进行统计推断，主要运用似然比的方法构造统计量；

第五部分是对本文提出的方法进行随机模拟和实证分析；

最后部分是对文章的总结和展望。

文章内容结构的流程框架如下所示：



1.5 本文的创新之处

①将 Guass-Seidel 型迭代算法引用到分层线性回归模型 EM 算法中，提高了算法的收敛速度。

②将 Guass-Seidel 型迭代算法推广到分层半参数模型中，体现了 Guass-Seidel 型迭代算法的普遍性。

③将具有嵌套结构的多元回归模型统计推断推广到分层线性回归嵌套模型的统计推断中，利用似然比方法构造统计量，对引入新的系数变量的必要性给出理论依据和判断准则。

2 算法理论

2. 1EM 算法

对于不完全数据包括截断数据，删失数据，既截断又删失数据以及缺失数据，其中缺失数据(missing data)可根据 Little 和 Rubin[4] (1987)将缺失数据分为三种类型：完全随机缺失(missing completely at random,MCAR)和随机缺失(missing at sandom,MAR)，以及非随机、不可忽略缺失(nonignorable)。MCAR 表示缺失数据部分具有随机性，与可观测数据部分是无关的，MAR 表示缺失数据部分是与观测数据部分相关的，同时在计算过程中，是不可以忽略缺失数据部分的，nonignorable 表示缺失数据依赖于可观测数据，若忽略会造成计算结果不精准。本文中 EM 算法主要针对随机缺失类型问题进行研究，在观测数据存在缺失的情况下加入随机缺失数据，形成完全数据，并对完全数据进行参数估计的过程。

EM 算法最初是由 Dempster[5]等(1997)提出，主要用于后验分布的众数（极大似然估计）的求解，EM 算法的迭代过程由 E 步（求期望）和 M 步（求极大似然值）两步组成，其中设Y为观测数据，Z为潜在数据，则(Y,Z)为完整数据。并以$p(\theta|Y)$表示参数θ关于观测数据Y的后验分布密度函数，称为观测后验分布，$p(\theta|Y,Z)$表示θ关于Y,Z的后验分布密度函数，称为添加后验分布，而且以$p(Z|\theta,Y)$表示潜在数据Z关于θ,Y的条件分布密度函数，EM 算法的目的是以计算$p(\theta|Y)$的众数，记θ^i第$i+1$次迭代开始时后验众数的估计值，则第$i+1$次迭代过程如下：

E 步：求$\log p(\theta|Y,Z)$关于Z的条件分布$p(Z|\theta,Y)$期望，即

$$Q(\theta|\theta^i,Y)=E_z(\log p(\theta|Y,Z)/\theta^i,Y)=\int\log(p(\theta|Y,Z))p(Z|\theta^i,Y)dZ$$

M 步：极大化$Q(\theta|\theta^i,Y)$，更新参数$\theta=\theta^i$，使得

$$Q(\theta^{(i+1)}|\theta^i,Y)=\max Q(\theta|\theta^i,Y)$$

如此形成了第$i+1$次迭代$\theta^i\to\theta^{i+1}$，重复上述 E 步和 M 步迭代，直至$|\theta^{i+1}-\theta^i|$或者$|Q(\theta^{i+1}|\theta^i,Y)-Q(\theta^i|\theta^i,Y)|$达到充分小的值时停止，得到最终参数

估计值。

引理 2.1[30] EM 算法在每一次迭代后均提高后验密度函数值，即

$$p(\theta^{(i+1)}|Y) \geqslant p(\theta^{(i)}|Y)$$

证明：由全概率公式$p(\theta, Z|Y) = p(Z|\theta, Y)p(\theta|Y) = p(\theta|Y, Z)p(Z|Y)$，将式子后面两项取对数得到

$$\log p(\theta|Y) = \log p(\theta|Y, Z) - \log p(Z|\theta, Y) + \log p(Z|Y)$$

现有参数估计值$\theta^{(i)}$，将上式对Z关于$p(Z|\theta^{(i)}, Y)$求期望，得到

$$\begin{aligned}\log p(\theta|Y) &= \int [\log p(\theta|Y, Z) - \log p(Z|\theta, Y) + \log p(Z|Y)]p(Z|\theta^{(i)}, Y)dZ \\ &\triangleq Q(\theta|\theta^{(i)}, Y) - H(\theta|\theta^{(i)}, Y) + K(\theta^{(i)}, Y)\end{aligned}$$

对应的

$$H(\theta|\theta^{(i)}, Y) = \int \log[p(Z|\theta, Y)]p(Z|\theta^{(i)}, Y)dZ$$

$$K(\theta^{(i)}, Y) = \int \log[p(Z|Y)]p(Z|\theta^{(i)}, Y)dZ$$

取$\log p(\theta|Y)$的第i次和第$i+1$次迭代相减，得到

$$\begin{aligned}&\log p(\theta^{(i+1)}|Y) - \log p(\theta^{(i)}|Y) \\ &= [Q(\theta^{(i+1)})|\theta^{(i)}, Y) - Q(\theta^{(i)}|\theta^{(i)}, Y)] - [H(\theta^{(i+1)}|\theta^{(i)}, Y) - H(\theta^{(i)}|\theta^{(i)}, Y)]\end{aligned}$$

由 *Jensen* 不等式可知

$$\begin{aligned}&E^{Z|\theta^{(i)},Y} \log(\tfrac{p(Z|\theta^{(i+1)}, Y)}{p(Z|\theta^{(i)}, Y)}) \\ &\leqslant \log\{E^{Z|\theta^{(i)}}(\tfrac{p(Z|\theta^{(i+1)}, Y)}{p(Z|\theta^{(i)}, Y)})\} \\ &= 0\end{aligned}$$

故$H(\theta^{(i+1)}|\theta^{(i)}, Y) - H(\theta^{(i)}|\theta^{(i)}, Y) \leqslant 0$，而$Q^{(i+1)}$是使$Q(\theta|\theta^{(i)}, Y)$达到最大值，显然可以得出

$$Q(\theta^{(i+1)}|\theta^{(i)}, Y) - Q(\theta^{(i)}|\theta^{(i)}, Y) \geqslant 0$$

证毕。

EM 算法的简单可行性和稳定性使得 EM 算法得到广泛的应用。

2. 2Guass-Seidel 型迭代算法

Guass-Seidel 型迭代算法是一种求解线性方程组常用的有效方法，其主要思想是在每次迭代中充分利用当前最新的迭代值，以加快收敛的速度，是数值代数中线性方程组求解的基本问题之一，其中许多工程与科学计算问题最后都可归结为求解线性方程组问题，例如利用 Guass-Seidel 型迭代实现迭代向量的循环传送，以尽量减少处理机间的通信次数。

例如对线性方程组$\sum_{j=0}^{n-1} a_{ij}x_i = b_i$，$i = 0,1,2,...,n-1$，那么 Guass-Seidel 型迭代算法公式为：

令$x^{(0)} = (x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)})^\top$作为初始变量。

$$x_i^{(k+1)} = \frac{1}{a_{ii}}(b_i - \sum_{j=0}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n-1} a_{ij}x_j^{(k)}) \quad i = 0,1,2,...,n-1$$

本文主要运用其每次迭代利用最新值的思想，把 Guass-Seidel 型迭代算法引入到分层线性模型 EM 算法中，使其下次迭代利用最新值，形成分层线性模型 Guass-Seidel 型迭代算法。

2. 3 收敛判断准则

在进行迭代运算过程中，迭代过程需依据判断准则停止，这里主要讨论基于分层线性模型的判断准则，以两层分层回归模型为例，迭代算法停止的依据是对数似然变化值足够小或者参数变化值足够小，我们可以选用完全数据对数似然变化值足够小或者迭代产生的最新值变化量足够小为判断准则，判断迭代算法是否收敛停止。可以设置似然函数的变化量或者参数最新估计值的变化量小于容忍度，而这个容忍值是可以自己设置的。

这里给出两层分层线性回归模型的一般形式

第一层模型：

$$Y_j = X_j\beta_j + r_j \quad , \quad r_j \sim N(0, \sigma^2)$$

第二层模型：

$$\beta_j = w_j\gamma + u_j \quad , \quad u_j \sim N(0, T)$$

将第二层模型带入到第一层模型中，得到更一般的形式，即混合效应模型

$Y_j = X_j w_j \gamma + X_j u_j + r_j$，这里关于模型的解释不详细介绍，具体可看下面章节模型部分内容，我们知道Y作为观测数据，θ_{rj}为缺失变量，那么(Y, θ_{rj})为完全数据，(θ_f, σ^2, T)作为待估计的未知参数值。

由于$Y_j|\theta_{rj} \sim N(A_{fj}\theta_f + A_{rj}\theta_{rj}, \sigma^2 I)$和$f(y;\theta_{rj}) = f(y|\theta_{rj})f(\theta_{rj})$，可以得到完整数据的似然函数为：

$$lnf(y;\theta_{rj}) \propto -nln(\sigma^2) - Jln(T) - \frac{1}{\sigma^2}\sum_{j=1}^{J}(y - A_{fj}\theta_f - A_{rj}\theta_{rj})^\top(y-$$

$$A_{fj}\theta_f - A_{rj}\theta_{rj}) - \sum_{j=1}^{J}\theta_{rj}^\top T^{-1}\theta_{rj}$$

记上式为$L(Y;\theta_{rj})$。

把迭代产生的新的参数估计值带入到判断准则：

$$\frac{L(\widehat{\theta}_f^{(i)}, \widehat{T}^{(i)}, \widehat{\sigma}^{2(i)}) - L(\widehat{\theta}_f^{(i-1)}, \widehat{T}^{(i-1)}, \widehat{\sigma}^{2(i-1)})}{L(\widehat{\theta}_f^{(i-1)}, \widehat{T}^{(i-1)}, \widehat{\sigma}^{2(i-1)})} < tolerance$$

对于 tolerance 的值介于 0 至 1 之间，一般情况下设置为 0.05，也可以根据文章内容而定。

3 分层模型 Guass-Seidel 型迭代算法

3.1 分层线性模型 EM 算法

在统计数据过程中，数据往往存在分层结构，例如研究高校间不同学生的学习情况，或用于研究国家经济发展的差异如何与成人教育程度相互作用，或研究临床药物的治疗方法的差异等，这些情况中存在嵌套问题的研究，分层线性回归模型给出了良好的模型结构。分层线性模型是由 Smith[31](1972)提出的，作为对线性模型的贝叶斯估计的重要贡献，同时对复杂的嵌套结构数据给出了通用的分层线性模型的形式。

这里以两层数据模型为例，给出两层结构的分层线性模型的具体形式，假设有(X,Y)的一组独立同分布观测值$\{(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)\}$，其中$Y_j$是第$j$组实数被解释变量，$X_j$是已知的第一层第$j$组的解释变量，$\beta_j$是未知的第$j$组的系数向量，满足第一层模型：

$$Y_j=X_j\beta_j+r_j \quad , \quad r_j\sim N\left(0,\sigma^2\right) \tag{3.1}$$

其中r_j是$i.i.d.$不可观测随机效应变量，假定与解释变量独立，并服从均值为 0，方差为σ^2的多元正态分布。

在第二层模型中，第一层模型中系数向量作为被解释变量，γ为固定效应向量，w_j为第二层已知的解释变量矩阵：

$$\beta_j=w_j\gamma+u_j \quad , \quad u_j\sim N\left(0,T\right) \tag{3.2}$$

其中u_j是第二层第j组的随机效应向量，假定与第二层解释变量和r_j独立，并服从均值向量为 0，协方差为T的多元正态分布。

将第二层模型带入第一层模型中，得到下列形式(Stephen 和 Raudenbush[7]1997)：

$$Y_j=X_jw_j\gamma+X_ju_j+r_j \quad , \quad j=1,...,J \tag{3.3}$$

下面给出比分层线性模型更具有一般形式的混合模型，因为它不要求第一层系数含有随机部分，因此混合模型的具体形式如下：

$$Y_j=A_{fj}\theta_f+A_{rj}\theta_{rj}+r_j，\ j=1,...,J \tag{3.4}$$

其中$A_{fj}=X_jw_j$，$\theta_f=\gamma$，$A_{rj}=X_j$，$\theta_{rj}=u_j$，A_{rj}作为具有随机效应X_j的子集，然而，在其他应用中，A_{rj}可能包含没有固定效应的变量。

线性混合模型是一类应用非常广泛的模型，可以用来分析处理纵向数据、组数据和面板数据等各类重复测量数据，突破了传统线性模型所要求的观测值是独立和等方差的条件限制，对观测值的协方差矩阵有了更灵活的设定，使在观测数据时有更好的方法。

相比于分层线性回归模型，线性混合效应模型针对多水平的随机效应部分给出了更方便且合理的假设。

本小节的主要内容来自于 Stephen 和 Raudenbush[7](1997)的分层线性模型的应用和数据分析方法，进行分层线性模型 EM 算法的具体推导。

在分层线性模型的参数估计问题中，将极大似然估计问题简单化，其迭代过程由 E 步和 M 步组成。

M（最大步）步：

Y_j为观察数据，θ_{rj}为缺失数据，那么(Y_j,θ_{rj})可作为完整数据，在上述分层线性回归模型中，可以看出θ_f，σ^2，T为待估计的未知参数。

从(3.4)式子可以推出：

$$Y_j-A_{rj}\theta_{rj}=A_{fj}\theta_f+r_j$$

根据普通最小二乘法可以得到：

$$\widehat{\theta}_f=(\sum A_{fj}^{\top}A_{fj})^{-1}\sum A_{fj}^{\top}(Y_j-A_{rj}\theta_{rj}) \tag{3.5}$$

根据r_j的方差可得到：

$$Var\left(r_j\right)=E\{\left(r_j-Er_j\right)\left(r_j-Er_j\right)^{\top}\}=E\left(r_jr_j^{\top}\right) \tag{3.6}$$

可以得出：

$$\begin{aligned}\widehat{\sigma}^2&=N^{-1}\sum\widehat{r}_j^{\top}\widehat{r}_j\\&=N^{-1}\sum\left(Y_j-A_{fj}\widehat{\theta}_f-A_{rj}\theta_{rj}\right)^{\top}\left(Y_j-A_{fj}\widehat{\theta}_f-A_{rj}\theta_{rj}\right)\end{aligned} \tag{3.7}$$

其中$N=\sum n_j$。

$$\widehat{T}=J^{-1}\sum\theta_{rj}\theta_{rj}^{\top} \tag{3.8}$$

将上述根据方差的定义和普通最小二乘法得到的参数估计值可以当作已知的参数初始值$\widehat{\theta}_f^{(0)}$，$\widehat{T}^{(0)}$，$\widehat{\sigma}^{2(0)}$。

E（期望步）步：

上述给出的参数初始值显式，并不是参数真实值，而是我们给出的形式上的参数估计初值，并不能直接进行计算，若我们想计算得到上述待估计参数的值，就必须知道下面完全数据充分统计量(CDSS)：

$$\sum Y_j^{\top} A_{rj}\theta_{rj}，\sum A_{fj}^{\top} A_{rj}\theta_{rj}，\sum \theta_{rj}^{\top} A_{rj}^{\top} A_{rj}\theta_{rj}，\sum \theta_{rj}\theta_{rj}^{\top}$$

虽然完全数据充分统计量没有办法直接被观测到，但是我们可以通过计算完全数据充分统计量的条件期望来得到完全数据充分统计量的估计值，以进行下面的迭代过程。

由于完全数据充分统计量无法直接观察得到，我们通过在给定参数的情况下，求完全数据充分统计量的条件期望来获得相关信息。

根据上述式子(3.4)可以得出完全数据的联合分布为：

$$\binom{Y_j}{\theta_{rj}} \sim N[\binom{A_{fj}\theta_f}{0}, \begin{pmatrix} A_{rj}TA_{rj}^{\top}+\sigma^2 I & A_{rj}T \\ TA_{rj}^{\top} & T \end{pmatrix}] \tag{3.9}$$

我们可以从上式(3.9)的联合分布得到被解释变量Y_j的期望和方差：

$$E\left(Y_j\right) = A_{fj}\theta_f，Var\left(Y_j\right) = A_{rj}TA_{rj}^{\top} + \sigma^2 I$$

其中在给定完全数据的情况下，缺失数据θ_{rj}的条件分布为：

$$\theta_{rj}|Y, \theta_f, T, \sigma^2 \sim N\left(\theta_{rj}^*, \sigma^2 C_j^{-1}\right)$$

其中相对应的：

$$\theta_{rj}^* = C_j^{-1} A_{rj}^{\top}\left(Y_j - A_{fj}\theta_f\right)$$

$$C_j = A_{rj}^{\top} A_{rj} + \sigma^2 T^{-1}$$

分层线性模型 EM 算法的迭代过程具体如下：

（1）估计完全数据充分统计量的条件期望：

$$E\left(\sum A_{fj}^{\top} A_{rj}\theta_{rj}|Y, \theta_f, \sigma^2, T\right) = \sum A_{fj}^{\top} A_{rj}\theta_{rj}^*$$
$$E\left(\sum \theta_{rj}\theta_{rj}^{\top}|Y, \theta_f, \sigma^2, T\right) = \sum \theta_{rj}^*\theta_{rj}^{*\top} + \sigma^2 \sum C_j^{-1}$$
$$E\left(\sum r_j^{\top} r_j|Y, \theta_f, \sigma^2, T\right) = \sum r_j^{*\top} r_j^* + \sigma^2 tr \sum C_j^{-1} A_{rj}^{\top} A_{rj}$$

其中，$r_j^* = Y_j - A_{fj}\theta_f - A_{rj}\theta_{rj}^*$。

（2）将估计得到的完全数据带入 M 步中式子(3.5)，(3.7)和(3.8)中，更新完全数据充分统计量的条件期望得到新的参数估计值，具体迭代过程如下：

关于$A_{fj}^\top A_{rj}\theta_{rj}$的条件期望更新为：

$$\begin{aligned}&E\left(A_{fj}^\top A_{rj}\theta_{rj}/Y,\widehat{\theta}_f^{(0)},\widehat{T}^{(0)},\widehat{\sigma}^{2^{(0)}}\right)\\&=\sum A_{fj}^\top A_{rj}C_j^{-1}A_{rj}^\top(Y_j - A_{fj}\widehat{\theta}_f^{(0)})\end{aligned}$$

关于$\theta_{rj}\theta_{rj}^\top$的条件期望为：

$$\begin{aligned}&E\left(\sum\theta_{rj}\theta_{rj}^\top/Y,\widehat{\theta}_f^{(0)},\widehat{T}^{(0)},\widehat{\sigma}^{2^{(0)}}\right)\\&=\sum\left(C_j^{-1}A_{rj}^\top\left(Y_j - A_{fj}\widehat{\theta}_f^{(0)}\right)\right)\left(C_j^{-1}A_{rj}^\top\left(Y_j - A_{fj}\widehat{\theta}_f^{(0)}\right)\right)^\top\\&\quad+\widehat{\sigma}^{2^{(0)}}\sum C_j^{-1}\end{aligned}$$

关于$r_j^\top r_j$的条件期望更新为：

$$\begin{aligned}&E\left(\sum r_j^\top r_j/Y,\widehat{\theta}_f^{(0)},\widehat{T}^{(0)},\widehat{\sigma}^{2^{(0)}}\right)\\&=\sum\left(Y_j - A_{fj}\widehat{\theta}_f^{(0)} - A_{rj}\left(C_j^{-1}A_{rj}^\top\left(Y_j - A_{fj}\widehat{\theta}_f^{(0)}\right)\right)\right)^\top\left(Y_j - A_{fj}\widehat{\theta}_f^{(0)}\right.\\&\quad\left.- A_{rj}\left(C_j^{-1}A_{rj}^\top\left(Y_j - A_{fj}\widehat{\theta}_f^{(0)}\right)\right)\right)+\widehat{\sigma}^{2^{(0)}}tr\sum C_j^{-1}A_{rj}^\top A_{rj}\end{aligned}$$

其中$C_j = A_{rj}^\top A_{rj} + \widehat{\sigma}^{2^{(0)}}\widehat{T}^{(0)^{-1}}$。

根据完全数据充分统计量的条件期望可以更新参数为：

$$\left(\widehat{\theta}_f^{(0)},\widehat{T}^{(0)},\widehat{\sigma}^{2^{(0)}}\right)\rightarrow\left(\widehat{\theta}_f^{(1)},\widehat{T}^{(1)},\widehat{\sigma}^{2^{(1)}}\right)$$

（3）将得到的这些新的参数估计值输入步骤（1）中，得到新的完全数据充分统计量的条件期望。

（4）继续直到对数似然变化足够小或者任何参数值的最大变化都足够小，例如判断准则为$|\widehat{\theta}_f^{(i+1)} - \widehat{\theta}_f^{(i)}| < tolerance$。

3. 2 分层线性模型 Guass-Seidel 迭代过程

结合分层线性模型 EM 算法的一般迭代过程和 Gauss-Seidel 型迭代的思想，文章在计算过程中，利用 EM 算法的基础上提出分层线性模型 Guass-Seidel 迭代方法，我们需要解决待估计的参数为θ_f，T,σ^2，分层线性模型 EM 算法是通过普

通最小二乘法和方差的定义去得到参数初始值$\widehat{\theta}_f^{(0)}$，$\widehat{T}^{(0)}$，$\widehat{\sigma}^{2^{(0)}}$，然后计算基于$Y$，$\theta_f$，$T,\sigma^2$的完全数据充分统计量的条件期望去更新得到$\widehat{\theta}_f^{(1)}$，$\widehat{T}^{(1)}$，$\widehat{\sigma}^{2^{(1)}}$，更新参数值直到对数似然变化变得足够小或者参数变化值足够小停止。

现在的迭代算法对应的是通过普通最小二乘法和方差的定义得到参数初始值$\widehat{\theta}_f^{(0)}$，$\widehat{T}^{(0)}$，$\widehat{\sigma}^{2^{(0)}}$，然后计算基于的完全数据充分统计量的条件期望去更新得到$\widehat{\theta}_f^{(1)}$，$\widehat{T}^{(1)}$，$\widehat{\sigma}^{2^{(1)}}$，在$\widehat{\theta}_f^{(1)},\widehat{T}^{(1)},\widehat{\sigma}^{2^{(1)}}$的基础上去计算$\widehat{\theta}_f^{(2)}\left(\widehat{\theta}_f^{(1)},\widehat{T}^{(1)},\widehat{\sigma}^{2^{(1)}}\right)$，在基于参数值$\widehat{\theta}_f^{(2)}\left(\widehat{\theta}_f^{(1)},\widehat{T}^{(1)},\widehat{\sigma}^{2^{(1)}}\right)$的基础上更新得到新的$\widehat{T}^{(2)}\left(\widehat{\theta}_f^{(2)},\widehat{\sigma}^{2^{(1)}},\widehat{T}^{(1)}\right)$,然后基于参数值$\widehat{T}^{(2)}\left(\widehat{\theta}_f^{(2)},\widehat{\sigma}^{2^{(1)}},\widehat{T}^{(1)}\right)$去更新得到参数值$\widehat{\sigma}^{2^{(2)}}\left(\widehat{\theta}_f^{(2)},\widehat{\sigma}^{2^{(1)}},\widehat{T}^{(2)}\right)$, 一直迭代下去，直至依据判断准则收敛。

具体迭代过程如下：

M（最大步）步：

根据普通最小二乘法和方差的定义计算得出分层线性模型中的待估计参数，计算过程与分层线性模型 EM 算法的一般迭代过程相同：

$$\widehat{\theta}_f^{(0)}=\left(\sum A_{fj}^{\top}A_{fj}\right)^{-1}\sum A_{fj}^{\top}\left(Y_j-A_{rj}\theta_{rj}\right)$$

$$\begin{aligned}\widehat{\sigma}^{2^{(0)}}&=N^{-1}\sum\widehat{r}_j^{\top}\widehat{r}_j\\&=N^{-1}\sum\left(Y_j-A_{fj}\widehat{\theta}_f^{(0)}-A_{rj}\theta_{rj}\right)^{\top}\left(Y_j-A_{fj}\widehat{\theta}_f^{(0)}-A_{rj}\theta_{rj}\right)\end{aligned}$$

$$\widehat{T}^{(0)}=J^{-1}\sum\theta_{rj}\theta_{rj}^{\top}$$

E（期望步）步：

将完全数据充分统计量的条件期望带入式子(3.5)，(3.7)和(3.8)得到新的估计参数值，即$\widehat{\theta}_f^{(1)},\widehat{T}^{(1)},\widehat{\sigma}^{2(1)}$，根据计算可得到：

更新得到的$\widehat{\theta}_f^{(1)}$估计值为:

$$\begin{aligned}\widehat{\theta}_f^{(1)}&=(\sum A_{fj}^{\top}A_{fj})^{-1}\sum(Y_j-A_{rj}\theta_{rj})\\&=\sum A_{fj}^{\top}A_{fj})^{-1}\sum A_{fj}^{\top}(Y_j-A_{rj}(A_{rj}^{\top}A_{rj}\\&\quad+\widehat{\sigma}^{2^{(0)}}\widehat{T}^{(0)^{-1}})^{-1}A_{rj}^{\top}(Y_j-A_{fj}\widehat{\theta}_f^{(0)}))\end{aligned}$$

更新得到的$\widehat{T}^{(1)}$估计值为:

$$\begin{aligned}\widehat{T}^{(1)} &= J^{-1}E\left(\sum \theta_{rj}\theta_{rj}^{\top}/Y,\theta_f,\sigma^2,T\right)\\ &= J^{-1}(\sum(C_j^{(0)})^{-1}A_{rj}^{\top}\left(Y_j-A_{fj}\widehat{\theta}_f^{(0)}\right)\left(\left(C_j^{(0)}\right)^{-1}A_{rj}^{\top}\left(Y_j-A_{fj}\widehat{\theta}_f^{(0)}\right)\right)^{\top}\\ &\quad+\widehat{\sigma}^{2^{(0)}}\sum(C_j^{(0)})^{-1})\end{aligned}$$

更新得到的$\widehat{\sigma}^{2^{(1)}}$估计值为:

$$\widehat{\sigma}^{2^{(1)}} = N^{-1}\left(\sum\left(r_j^{*^{(0)}}\right)^{\top}\left(r_j^{*^{(0)}}\right)+\widehat{\sigma}^{2^{(0)}}tr\sum\left(C_j^{(0)}\right)^{-1}A_{rj}^{\top}A_{rj}\right)$$

其中$C_j^{(0)} = A_{rj}^{\top}A_{rj}+\widehat{\sigma}^{2^{(0)}}\widehat{T}^{(0)^{-1}}$。

其中$r_j^{*^{(0)}} = Y_j-A_{fj}\widehat{\theta}_f^{(0)}-A_{rj}\left(C_j^{(0)}\right)^{-1}A_{rj}^{\top}\left(Y_j-A_{fj}\widehat{\theta}_f^{(0)}\right)$。

分层线性模型 Guass-Seidel 型迭代过程如下：

（1）初始化参数值$(\widehat{\theta}_f,\widehat{T},\widehat{\sigma}^2) = (\widehat{\theta}_f^{(0)},\widehat{T}^{(0)},\widehat{\sigma}^{2^{(0)}})$；

（2）估计完全数据充分统计量的条件期望更新参数值得到$\widehat{\theta}_f^{(1)}$，$\widehat{T}^{(1)}$，$\widehat{\sigma}^{2^{(1)}}$；

（3）根据完全数据充分统计量的条件期望得到的最新参数值带入到 M 步得到：

基于$\widehat{\theta}_f^{(1)}$，$\widehat{T}^{(1)}$，$\widehat{\sigma}^{2^{(1)}}$的基础上$\widehat{\theta}_f^{(2)}$的最新估计值：

$$\begin{aligned}&\widehat{\theta}_f^{2}(\widehat{\theta}_f^{(1)},\widehat{T}^{(1)},\widehat{\sigma}^{2^{(1)}})\\ &=(\sum A_{fj}^{\top}A_{fj})^{-1}\sum A_{fj}^{T}(Y_j\\ &\quad-A_{rj}(A_{rj}^{\top}A_{rj}+\widehat{\sigma}^{2^{(1)}}\widehat{T}^{(1)^{-1}})^{-1}A_{rj}^{\top}(Y_j-A_{fj}\widehat{\theta}_f^{(1)})\end{aligned}$$

基于$\widehat{\theta}_f^{(2)}$，$\widehat{T}^{(1)}$，$\widehat{\sigma}^{2^{(1)}}$的基础上$\widehat{T}^{(2)}$的最新估计值：

$$\begin{aligned}&\widehat{T}^{(2)}\left(\widehat{\theta}_f^{(2)},\widehat{T}^{(1)},\widehat{\sigma}^{2^{(1)}}\right)\\ &=J^{-1}\left[\sum\left(C_j^{(1)}\right)^{-1}A_{rj}^{\top}\left(Y_j-A_{fj}\widehat{\theta}_f^{(2)}\right)\left(\left(C_j^{(1)}\right)^{-1}A_{rj}^{\top}(Y_j\right.\right.\\ &\quad\left.-A_{fj}\widehat{\theta}_f^{(2)^{\top}}+\widehat{\sigma}^{2^{(1)}}\left(C_j^{(1)}\right)^{-1}\right]\end{aligned}$$

基于$\widehat{\theta}_f^{(2)}$，$\widehat{T}^{(2)}$，$\widehat{\sigma}^{2^{(1)}}$的基础上$\widehat{\sigma}^{2^{(2)}}$的最新估计值：

$$\begin{aligned}&\widehat{\sigma}^{2^{(2)}}\left(\widehat{\theta}_f^{(2)},\widehat{T}^{(2)},\widehat{\sigma}^{2^{(1)}}\right)\\&=N^{-1}\left(\sum\left(r_j^{*^{(1)}}\right)^{\top}\left(r_j^{*^{(1)}}\right)+\widehat{\sigma}^{2^{(1)}}tr\sum\left(C_j^{(1)}\right)^{-1}A_{rj}^{\top}A_{rj}\right)\end{aligned}$$

对应的$C_j^{(1)}=A_{rj}^{\top}A_{rj}+\widehat{\sigma}^{2^{(1)}}\widehat{T}^{-1^{(2)}}$。

对应的$r_j^{*^{(1)}}=Y_j-A_{fj}\widehat{\theta}_f^{(2)}-A_{rj}\left(C_j^{(1)}\right)^{-1}A_{rj}^{\top}\left(Y_j-A_{fj}\widehat{\theta}_f^{(2)}\right)$。

（4）重复上述过程，迭代产生$\widehat{\theta}_f^{(3)}(\widehat{\theta}_f^{(2)},\widehat{T}^{(2)},\widehat{\sigma}^{2^{(2)}})$，$\widehat{T}^{(3)}(\widehat{\theta}_f^{(3)},\widehat{T}^{(2)},\widehat{\sigma}^{2^{(2)}})$，$\widehat{\sigma}^{2^{(3)}}(\widehat{\theta}_f^{(3)},\widehat{T}^{(3)},\widehat{\sigma}^{2^{(2)}})$，⋯ ，一直迭代下去，直到依据判断准则收敛停止，例如$|\widehat{T}^{i+1}-\widehat{T}^{i}|<tolerance$。

3. 3 Guass-Seidel 型迭代算法在分层半参数模型的扩展

根据上述提出的分层线性回归模型 Guass-Seidel 型迭代算法的迭代求参数过程，下面提出分层模型 Guass-Seidel 型迭代算法的一种模型扩展，即基于缺失数据情况下的分层半参数模型的 Guass-Seidel 型迭代算法。

这里考虑模型第一层为半参数模型，且第一层模型是基于同方差的情况下，同时当$E(\varepsilon_{ij}^2)=\sigma_i^2(X_{ij})$，它允许组内出现异方差情况，这里不具体讨论。

在线性模型中，通常假设$\{(X_{ij},W_{ij},Y_{ij})\}$，$i=1,...,n$，$j=1,...,n_i$来自总体$(X,W,Y)$, 其中$Y_{ij}$为第$i$个单元的第$j$个个体，$X_{ij}$为已知的$d\times 1$维第一层的解释变量，$W_i$为$d\times f$维第二层的解释变量矩阵。这里，假设分层模型第一层考虑半参数模型(田茂再[15]2015)：

$$Y_{ij}=m(X_{ij};\beta_i)+\varepsilon_{ij}，i=1,...,n，j=1,...,n_i。$$

其中$m(\cdot)$为未知函数，β_i是不可以直接观测到的随机变量，ε_{ij}是不可观测的随机误差变量，服从均值为 0，方差为$\sigma^2 I$的多元正态分布。用β_i的好处包括维数的减少和高水平变量效应的识别，大量的观测变量可以更好的理解数据。

假设$m(\cdot)$在x处有一阶连续偏导，对于x周围的样本点通过泰勒线性展开估计非参数部分$m(X_{ij};\beta_i)$，如$m(X_{ij};\beta_i)\approx m(x;\beta_i)+(X_{ij}-x)^{\top}\bigtriangledown m(x;\beta_i)$，从而使

$$m(X_{ij},\beta_i)\approx \widetilde{X}_{ij}^{\top}\theta_i$$

其中$\widetilde{X}_{ij}=(1,(X_{ij}-x)^{\top})^{\top}$，$\theta_i=(m(x;\beta_i),\triangledown m(x;\beta_i)^{\top})^{\top}=(\theta_{i0},...,\theta_{id})^{\top}$。

第二层模型：

$$\theta_i=W_i\gamma(x)+U_i$$

其中W_i表示第二层解释变量矩阵，$\gamma(x)$为固定效应函数，U_i表示第二层的随机向量，服从均值为 0，协方差矩阵为T。

将第二层模型带入到第一层模型中可以得出

$$Y_{ij}=\widetilde{X}_{ij}^{\top}W_i\gamma(x)+\widetilde{X}_{ij}^{\top}U_i+\varepsilon_{ij}$$

根据上述模型，我们可以清楚的看出现阶段带估计的参数为$\gamma(x)$，σ^2，T，这里我们令$A_{fj}=\widetilde{X}_{ij}^{\top}W_i$，$A_{rj}=\widetilde{X}_{ij}^{\top}$，这时具有更一般的混合效应模型的形式：

$$Y_{ij}=A_{fj}\gamma(x)+A_{rj}U_i+\varepsilon_{ij}$$

这时，对应的

$$\widehat{\gamma}(x)=(\sum A_{fj}^{\top}A_{fj})^{-1}\sum A_{fj}^{\top}(Y_{ij}-A_{rj}U_i)$$

可以得出完全数据的联合分布为：

$$\binom{Y_j}{U_i}\sim N[\binom{A_{fj}\gamma(x)}{0},\begin{pmatrix}A_{rj}TA_{rj}^{\top}+\sigma^2 I & A_{rj}T\\ TA_{rj}^{\top} & T\end{pmatrix}]$$

给定完全数据下的缺失数据U_i的条件分布为：

$$U_i|Y,\gamma(x),T,\sigma^2\sim N\left(U_i^*,\sigma^2C_j^{-1}\right)$$

其中相对应的：

$$U_i^*=C_j^{-1}A_{rj}^{\top}\left(Y_{ij}-A_{fj}\gamma(x)\right)$$

$$C_j=A_{rj}^{\top}A_{rj}+\sigma^2T^{-1}$$

对于σ^2，$\gamma(x)$和T的估计，使用 EM 算法。

M 步（最大步）：根据σ^2，T的最大似然估计得到

$$\widehat{T}=\frac{1}{n}\sum_{i=1}^{n}U_iU_i^{\top}$$

$$\widehat{\sigma}^2=\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-A_{fj}\widehat{\gamma}(x)-A_{rj}U_i)^2$$

E 步（期望步）：根据σ^2，$\gamma(x)$和T的条件期望为

$$E(\sum A_{fj}^{\top}A_{rj}U_i|Y,\gamma(x),\sigma^2,T)=\sum A_{fj}^{\top}A_{rj}U_i^*$$

$$E(\widehat{T}|Y,\gamma(x),\sigma^2,T)=\tfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}(U_i^*U_i^{*\top}+T_i^*)$$

$$E(\widehat{\sigma}^2|Y,\gamma(x),\sigma^2,T)=\tfrac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-A_{fj}\widehat{\gamma}(x)-A_{rj}U_i^*)^2+tr(A_{rj}^{\top}T_i^*A_{rj})$$

其中$T_i^*=\widehat{\sigma}^2(A_{rj}^{\top}A_{rj}+\widehat{\sigma}^2\widehat{T}^{-1})^{-1}$，其中相对应的：$U_i^*=C_j^{-1}A_{rj}^{\top}(Y_{ij}-A_{fj}\widehat{\gamma}(x))$。按照 EM 算法迭代进行下去，直到依据所有参数变化量小于给定的容忍度这一收敛判断准则迭代算法停止。

下面给出引入 Guass-Seidel 型迭代算法在分层半参数模型的一般推广过程，给出具体的参数求解过程，体现迭代算法的普遍性。

基于 Guass-Seidel 型迭代的分层半参数求解参数过程如下：

（1）初始化参数值$(\widehat{\gamma}(x)^{(0)},\widehat{\sigma}^{2^{(0)}},\widehat{T}^{(0)})$，即为

$$\widehat{\gamma}(x)^{(0)}=(\sum A_{fj}^{\top}A_{fj})^{-1}\sum A_{fj}^{\top}(Y_{ij}-A_{rj}U_i)$$

$$\widehat{\sigma}^{2^{(0)}}=\tfrac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(0)}-A_{rj}U_i)^2$$

$$\widehat{T}^{(0)}=\tfrac{1}{n}\sum_{i=1}^{n}U_iU_i^{\top}$$

（2）根据期望步的条件期望更新参数可以得出

$$\widehat{\gamma}(x)^{(1)}=(\sum A_{fj}^{\top}A_{fj})^{-1}\sum A_{fj}^{\top}(Y_{ij}-A_{rj}C_j^{-1^{(0)}}A_{rj}^{\top}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(0)}))$$

$$\widehat{T}^{(1)}=\tfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}(C_j^{-1^{(0)}}A_{rj}^{\top}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(0)}))^2+\widehat{\sigma}^{2^{(0)}}C_j^{-1^{(0)}}$$

$$\begin{aligned}\widehat{\sigma}^{2^{(1)}}&=\frac{1}{n_i}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(0)}-A_{rj}C_j^{-1^{(0)}}A_{rj}^{\top}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(0)}))^2\\&\quad+tr(A_{rj}^{\top}\widehat{\sigma}^{2^{(0)}}(A_{rj}^{\top}+\widehat{\sigma}^{2^{(0)}}\widehat{T}^{-1^{(0)}})^{-1}A_{rj})\\&=\frac{1}{n_i}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(0)}-A_{rj}U_i^{*^{(0)}})^2+tr(A_{rj}^{\top}T_i^{*^{(0)}}A_{rj})\end{aligned}$$

对应到上述式子，$C_j^{(0)}=A_{rj}^{\top}A_{rj}+\widehat{\sigma}^{2^{(0)}}\widehat{T}^{-1^{(0)}},U_i^{*^{(0)}}=C_j^{-1^{(0)}}A_{rj}^{\top}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(0)})$和$T_i^{*^{(0)}}=\widehat{\sigma}^{2^{(0)}}(A_{rj}^{\top}+\widehat{\sigma}^{2^{(0)}}\widehat{T}^{-1^{(0)}})^{-1}$。

（3）将条件期望更新得到的参数带入到 M 步中基于$\widehat{\gamma}(x)^{(1)}$，$\widehat{T}^{(1)}$，$\widehat{\sigma}^{2^{(1)}}$情况下可以得到$\widehat{\gamma}(x)^{(2)}=(\sum A_{fj}^{\top}A_{fj})^{-1}\sum A_{fj}^{\top}(Y_{ij}-A_{rj}C_j^{-1^{(1)}}A_{rj}^{\top}(Y_{ij}-A_{fj}\widehat{\gamma}(x)^{(1)}))$；

（4）依据迭代算法反复迭代下去，可以得出$\widehat{T}^{(2)}\left(\widehat{\gamma}(x)^{(2)},\widehat{T}^{(1)},\widehat{\sigma}^{2^{(1)}}\right)$，

$\widehat{\sigma}^{2^{(2)}}\left(\widehat{\gamma}(x)^{(2)}, \widehat{T}^{(2)}, \widehat{\sigma}^{2^{(1)}}\right)$，⋯，直到所有参数的绝对变化小于一个预先给定充分小的数为止。

对于上述分层模型是基于同方差情况下的，由于ε_{ij}是不可观测的随机误差变量，它允许组内出现异方差，即方差为$E(\varepsilon_{ij}^2)=\sigma_i^2(X_{ij})$，这时对固定效应$\gamma(x)$的估计使用普通最小二乘方法的推理是失效的，这时可能需要考虑加权普通二乘法。如果假设第一层是异方差的情况下，即$E(\varepsilon_{ij}^2)=\sigma_i^2(X_{ij})$有平滑函数$\sigma_i^2(\cdot)$，可以考虑在 x 点局部方差估计作为 $\sigma_i^2(x)$ 的一个初始化估计，即 $\widehat{\sigma}_i{}^2(x)=(\sum_{j=j_1}^{j_2} w_j Y_{ij})^2$，其中 $j_1=-[m/2]$，$j_2=[m/2-1/4]$，$\sum_{j=j_1}^{j_2} w_j=0$ 和 $\sum_{j=j_1}^{j_2} w_j^2=1$，而$m\geq 2$是一个固定的整数，参见 Stadtmuller[32](1987)。

4 分层模型统计推断

统计推断是研究如何利用样本数据来推断总体特征的统计方法，例如，要了解一个地区的人口特征不可能对每个人的特征进行测量，则需要通过部分样本来推断得到一个地区的特征。统计推断包括参数估计和总体参数的假设检验，上章节详细介绍了分层线性模型的参数估计过程，本章节主要介绍假设检验内容，求解检验过程的方法包括似然比检验，Bayes 检验和并-交检验与交-并检验，对检验的评价方法主要包括错误概率与功效函数，最大功效检验，并-交检验与交-并检验的真实水平，P 值，损失函数最优化，本章节主要构造分层线性模型似然比统计量来检验模型系数的拟合问题。

4.1 对第一层模型系数的推断

针对具有嵌套结构的分层线性回归模型进行统计推断，若我们只考虑分层线性模型的第一层系数的回归诊断，那分层线性模型就可直接理解为多元线性回归模型的回归诊断。为了更好的进行第一层模型统计诊断，现在将第一层模型进行形式上的变换，其实际意义并无影响。

我们考虑数据存在随机线性模型的一般形式，线性模型意味着假定因变量y和自变量x之间的关系可以用线性关系来近似(吴喜之[22]2016)：

$$Y_i = X_i^{\top}\beta + \varepsilon_i,\ \ i = 1, ..., n$$

其中Y_i为$n\times 1$的观测向量，X_i为$n\times p$已知的向量矩阵，β为待估计的未知参数，ε_i是模型所无法描述的随机误差项。通常情况下，随机误差ε_i满足 3 个假设：（1）$E\left(\varepsilon_i\right) = 0$；（2）$Var\left(\varepsilon_i\right) = \sigma^2$；（3）$Cov\left(\varepsilon_i, \varepsilon_j\right) = 0$，$i \neq j$。

通常情况下，人们把未知的σ^2假设为相等，若这一假定不成立，则称线性回归模型存在异方差性，在存在异方差性情况下用传统的最小二乘法估计模型参数，得到的参数估计量不是有效估计量，这里不再具体介绍异方差情况下的回归诊断情况，下面线性回归诊断是基于同方差情况下介绍的。

对于线性回归模型的待估计参数β常用的估计参数的方法是普通最小二乘法，其目的是使得$\varepsilon = y - x\beta$达到最小，即$S\left(\beta\right) = \left(y - x\beta\right)^2$达到最小。即对未

知参数β求偏导数，令函数$S\left(\beta\right)$为零，可以得到：

$$\widehat{\beta}=\left(X^{\top}X\right)^{-1}X^{\top}y$$

如果一个模型中的自变量为另一个模型的自变量的子集或者子集的线性组合，则称两个模型为嵌套模型(谢宇[23]2013)。一个模型子集或子集的线性组合的模型称为限制性模型，对应的另一个模型称为非限制性模型，限制性模型嵌套在非限制性模型中。

对多元线性回归模型的系数提出假设如下：

$$H0\colon\ Y=X_1\beta_1+X_2\beta_2+\varepsilon\quad vs\quad H1\colon\ Y=X_1\beta_1+X_2\beta_2+X_3\beta_3+\varepsilon$$

对应的检验统计量：

$$F_j=\frac{\left(SSE_{H0}-SSE_{H1}\right)/\left(k-g\right)}{SSE_{H1}/\left(n-k-1\right)}$$

其中SSE_{H0}为原假设H_0对应的残差平方和，SSE_{H1}为备择假设H_1对应的残差平方和，也称为限制性模型和非限制模型的残差平方和(谢宇[23]2013)。这里k对应备择假设模型所包含的回归系数的数量,则$n-k-1$对应备择假设残差平方和的自由度。其中$k-g$这个自由度增量是备择假设与原假设对应模型之间回归系数个数的差值。

对于给定的显著性水平α，检验的拒绝域为$F_j>F_{\alpha}\left(1,n-k-1\right)$。由于原假设去掉部分自变量，所以理论上原假设对应的残差平方和不小于备择假设的残差平方和。

由于这里原假设与备择假设对应模型之间只差一个参数，所以也可以使用t检验统计量$t_{df}^2=F_{1,df}\sim t\left(n-k-1\right)$，对于给定的显著性水平$\alpha$，检验的拒绝域为$|t_j|>t_{\alpha/2}\left(n-k-1\right)$。其中$F$统计量的第一自由度为 1，这时既可以使用$F$统计量也可以使用$t$统计量。

对于上述嵌套结构的线性回归模型，同时也可以使用判定系数增量来解释回归模型拟合优度的问题，其具体为非限制性模型的判定系数减去限制性模型的判定系数。

4.2 对第二层模型系数的推断

具有嵌套结构的模型是普遍存在的，同时利用似然比构造统计量判断模型的拟合问题也是十分常见的，例如关于对二分因变量进行嵌套模型分析的统计方法(谢宇[23]2013)，其目的在于估计和预测成功或失败的概率是否受到协变量的影响。针对存在嵌套关系的二分因变量，常通过对数似然比检验来判断模型的拟合优度问题，即具体为两个嵌套模型之间的对数似然比之差构造统计量，其统计量服从χ^2分布，相应为的统计量形式为：

$$\Delta G^2 = G_r^2 - G_u^2$$

其中G_r^2表示约束模型的对数似然比，G_u^2为无约束模型的对数似然比，则二分因变量嵌套模型服从的χ^2分布对应的自由度为无约束模型的残差自由度与约束模型的残差自由度之差。

针对多层线性回归模型的情况，模型中固定效应和随机效应部分的参数估计结果都有相应的检验方法。对于固定部分的显著性检验，可以用参数估计值除以标准误，即对应的$\gamma/se\left(\gamma\right)$进行，此方法称为 *Wald* 检验。对于随机部分的参数检验，可以近似采用估计得到的方差与标准误的比值，用 Z 检验进行检验，但由于 Z 检验的前提条件往往不易满足，建议使用卡方分布(刘红云[26]2005)。

同时给出了差异统计量的定义，对两个具有嵌套关系模型拟合程度进行检验，主要利用模型整体拟合差异统计量这一统计指标，两个模型差异统计量的差值理论上服从卡方分布，其对应自由度为两个模型参数个数的差(刘红云[26]2005)。

结合上述具有嵌套结构的多元线性回归模型的检验，二分因变量的嵌套模型检验，以及对分层线性模型中固定部分和随机部分参数估计结果的检验，以及使用模型整体拟合的差异统计量来检验嵌套模型的方法，扩展到分层线性回归模型中，下面主要内容是得出分层线性模型的似然函数。

将第二层模型带入到第一层模型中，Y_i具有线性混合模型形式，已知第一层随机误差$r_i \sim N\left(0, \sigma^2\right)$和第二层随机误差$u_i \sim N\left(0, T\right)$，所以线性混合模型$Y_i$服从$y \sim N\left(xwr, V\right)$，相应的，$V = xTx^\top + \sigma^2 I_n$。从而给出分层线性模型的似然

函数为：

$$
\begin{aligned}
L(y) &= \prod_{j=1}^{J}[(2\pi)^{-\frac{n}{2}}|V|^{\frac{1}{2}}\exp\{-\frac{1}{2}(Y-Xwr)^{\top}V^{-1}(Y-Xwr)\ \}] \\
&= \prod_{j=1}^{J}[(2\pi)^{-\frac{n}{2}}\left|x_j\tau x_j^{\top}+\sigma^2 I_n\right|^{\frac{1}{2}}\exp\{-\frac{1}{2}(Y-Xwr)^{\top}\left|x_j\tau x_j^{\top}\right. \\
&\quad \left.+\sigma^2 I_n\right|^{-1}(Y-Xwr)\ \}] \\
&= (2\pi)^{-\frac{n^2}{2}}\left|x_j\tau x_j^{\top}+\sigma^2 I_n\right|^{\frac{n}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(Y-Xwr)^{\top}\left|x_j\tau x_j^{\top}\right. \\
&\quad \left.+\sigma^2 I_n\right|^{-1}(Y-Xwr)\ \}
\end{aligned}
$$

其中，$T=diag(\tau,...,\tau)$，且$j=1,...,J$。

根据上述内容，下面给出具有嵌套结构的分层线性模型的一般假定情况：

零假设情况下：

第一层：$Y_1=X\beta_1+\varepsilon_1$ ，$\varepsilon_1\sim N(0,\sigma_1^2)$

第二层： $\beta_1\sim N(w_1 r,T_1)$

备择假设情况下：

第一层：$Y_2=X_1\beta_1+X_2\beta_2+\varepsilon_2$ ，$\varepsilon_2\sim N(0,\sigma_2^2)$

第二层： $\binom{\beta_1}{\beta_2}\sim N\left(\binom{w_1 r}{w_2 r},T_2\right)$

假设$L_n(\theta\mid y)$是参数θ的似然函数，其中$y=(y_1,...,y_n)'$是一个样本容量为n的样本，参数θ的参数空间为Ω，检验问题为$H0:\theta\in\Omega_0\ \ vs\ \ H_1:\theta\notin\Omega_0$，则统计量定义似然比(吴密霞[25]2013)为：

$$LR_n(y)=\frac{sup_{\theta\in\Omega_0}L_n(\theta|y)}{sup_{\theta\in\Omega}L_n(\theta|y)}$$

基于似然比统计量的定义，这里我们构造分层线性回归嵌套模型检验的似然比，原假设为限制性分层线性回归模型，备择假设为非限制性分层线性回归模型，从而构造具有嵌套结构的分层线性回归模型的似然比检验统计量。

零假设情况下似然函数：

$$L(H_0)=\prod_{i=1}^{n}[(2\pi)^{-\frac{n}{2}}|V_1|^{\frac{1}{2}}\exp\{-\tfrac{1}{2}(Y_{1i}-X_{1i}w_1 r)^{\top}V_1^{-1}(Y_{1i}-X_{1i}w_1 r)\}\ ]$$

其中 $V_1=x_1 T_1 x_1^{\top}+\sigma_1^2 I_n$。

备择假设情况下似然函数：

$$L(H_1)=\prod_{i=1}^{n}[(2\pi)^{-\frac{n}{2}}\left|V_2\right|^{\frac{1}{2}}\exp\{-\frac{1}{2}(Y_{2i}-X_{1i}w_1r- \\ X_{2i}w_2r)^{\top}V_2^{-1}(Y_{2i}-X_{1i}w_1r-X_{2i}w_2r)\}]$$

其中$V_2=x_1T_1x_1^{\top}+x_2T_2x_2^{\top}+\sigma_2^2I_n$。

则构造统计量为：

$$\begin{aligned}\frac{L\left(H_0\right)}{L\left(H_1\right)} &= \frac{\prod_{i=1}^{n}[(2\pi)^{-\frac{n}{2}}\left|V_1\right|^{\frac{1}{2}}\exp\{-\frac{1}{2}\left(Y_{1i}-X_{1i}w_1r\right)^{\top}}{\prod_{i=1}^{n}[(2\pi)^{-\frac{n}{2}}\left|V_2\right|^{\frac{1}{2}}\exp\{-\frac{1}{2}\left(Y_{2i}-X_{1i}w_1r-X_{2i}w_2r\right)} \cdot \\ &\quad \frac{V_1^{-1}(Y_{1i}-X_{1i}w_1r)\}]}{V_2^{-1}(Y_{2i}-X_{1i}w_1r-X_{2i}w_2r)\}]} \\ &= \frac{\left|V_1\right|^{\frac{1}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}\left(Y_{1i}-X_{1i}w_1r\right)^{\top}}{\left|V_2\right|^{\frac{1}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}\left(Y_{2i}-X_{1i}w_1r-X_{2i}w_2r\right)^{\top}} \cdot \\ &\quad \frac{V_1^{-1}(Y_{1i}-X_{1i}w_1r)\}}{V_2^{-1}(Y_{2i}-X_{1i}w_1r-X_{2i}w_2r)\}}\end{aligned}$$

其中分子表示备择假设的似然函数最大值，分母表示原假设下的似然函数最大值，如果统计量的值很大，说明原假设情况的可能性比备择假设情况下的可能性要小，于是，我们有理由认为原假设不成立。

对于多层嵌套线性回归模型，构造统计量为：

$$\Lambda\left(x_1,...,x_n\right)=-2\log\frac{L(H_0)}{L(H_1)}$$

引理 4.1[36] 关于检验$H_0:\theta=\theta_0\ \ vs\ \ H_1:\theta\neq\theta_0$，设$X_1,...,X_n$是$iid$的$f(x|\theta)$，$\widehat{\theta}$是$\theta$的 MLE, 并且$f(x|\theta)$满足：（1）我们观测到$X_1,...,X_n$，$X_i\sim f(x|\theta)$是$iid$的；（2）参数是可识别的，即如果$\theta\neq\theta'$，则$f(x|\theta)\neq f(x|\theta')$；（3）各个密度$f(x|\theta)$有共同支撑集，并且$f(x|\theta)$关于$\theta$可导；（4）参数空间$\Omega$包含一个开集$\omega$，以真参数值$\theta_0$为该开集的一个内点；（5）对于每个$x\sim\mathcal{X}$，密度$f(x|\theta)$关于$\theta$是三阶可导的，其三阶导数是$\theta$的连续函数，并且$\int f(x|\theta)dx$可以在积分号下微分三次；（6）对任何$\theta_0\sim\Omega$，存在一个正数$c$和一个函数$M(X)$使得

$$\left|\frac{\partial^3}{\partial\theta^3}\log f(x|\theta)\right|\leqslant M(x)，对于所有x\sim\mathcal{X}，\theta_0-c<\theta<\theta_0+c，$$

以及$E_{\theta_0}|M(X)|<\infty$。

则在H_0之下，当$n \to \infty$，

$$-2\log\lambda(X) \xrightarrow{\mathrm{L}} \chi_1^2$$

其中χ_1^2是一个自由度为 1 的χ^2分布随机变量。

证明：首先在$\hat{\theta}$的邻域展开$\log L(\theta|x) = l(\theta|x)$为 Taylor 级数，有

$$l(\theta|x) = l(\hat{\theta}|x) + l'(\hat{\theta}|x)(\theta - \hat{\theta}) + l''(\hat{\theta}|x)\frac{(\theta - \hat{\theta})^2}{2} + \ldots$$

现在把$l(\theta_0|x)$的展开式带入$-2\log\lambda(x) = -2l(\theta_0|x) + 2l(\hat{\theta}|x)$中，得到

$$-2\log\lambda(x) \approx \frac{(\theta - \hat{\theta})^2}{-l''(\hat{\theta}|x)}$$

现在用到$l'(\hat{\theta}|x) = 0$这个事实，因为分母就是观测信息数$\hat{I}_n(\hat{\theta})$并且$\hat{I}_n(\hat{\theta}) \to I(\theta_0)$，于是根据 MLE 的渐近有效性和 Slutsky 定理可推断出$-2\log\lambda(X) \xrightarrow{\mathrm{L}} \chi_1^2$。

该统计量服从χ^2分布，其自由度等于备择假设参数的个数减去原假设中参数的个数，对于本文中给出的具有嵌套结构的分层线性回归模型零假设与备择假设情况，具体的χ^2分布对应的自由度为 2。

在显著性水平α下，其拒绝域为：$W = \{\Lambda\left(x_1, \ldots, x_n\right) > \chi_2^2 = c\}$，如果落入拒绝域中，说明统计诊断不显著，则拒绝原假设，接受备择假设非限制性分层线性模型。

5 数值模拟和案例分析

5.1 随机模拟

对于分层线性回归模型，假设第一层解释变量X_j为$2 \times n_i$维，且第一列元素为 1，第二列元素为服从正态分布$N(0,2)$，同时第一层随机效应向量r_j服从正态分布$N(0,\sigma^2)$，令$\sigma^2=1$，令第二层解释变量w_i中元素服从正态分布$N(0,1)$，第二层固定效应向量为$\gamma=[2,3,1,1]^\top$，对应协方差的真值为$(0.5,0,0,0.5)$，得到被解释变量Y_j为$n_iJ\times 1$维向量，其中$n_i=20$，$J=40$。

下面给出迭代算法初始值用于分层线性回归模型 EM 算法参数估计和分层线性回归模型 Guass-Seidel 型迭代算法参数估计，令$\widehat{\sigma}_0^2=0.2$，固定效应向量初始值为$\widehat{\gamma}_0=[1,1,1,1]^\top$，协方差迭代初始值$\widehat{\tau}_{00_0}=0.2$，$\widehat{\tau}_{01_0}=0$，$\widehat{\tau}_{10_0}=0$，$\widehat{\tau}_{11_0}=0.2$，根据设置迭代次数为 5，50，500，1000 来观测 EM 算法和 Guass-Seidel 型迭代算法得到的参数估计值，进而得到有效信息。

表 5.1 分层线性模型参数估计结果（EM 算法）

迭代次数	σ^2	τ_{00}	τ_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	1.050	4.547	0.537	1.111	1.180	1.006	0.991
50	1.050	1.122	0.532	1.836	2.359	1.041	0.937
500	1.050	0.536	0.531	2.260	3.046	1.059	0.914
1000	1.050	0.536	0.531	2.260	3.046	1.059	0.914

根据模拟计算可以直观的看出依据迭代次数 5，50，500，1000 的变化，分层线性回归模型 EM 算法得到的参数估计值，可以看出有明显的收敛趋势，在迭代 500 次后，参数的估计收敛到固定值，对应的参数估计值为$\widehat{\sigma}^2=1.050$，协方差估计值对应为$[\widehat{\tau}_{00},\widehat{\tau}_{01},\widehat{\tau}_{10},\widehat{\tau}_{11}]=[0.536,0.023,0.023,0.531]$，固定效应的估计值

对应为$[\widehat{\gamma}_{00},\widehat{\gamma}_{01},\widehat{\gamma}_{10},\widehat{\gamma}_{11}]=[2.260,3.046,1.059,0.914]^\top$。

表 5.2 分层线性模型参数估计结果（Guass-Seidel 算法）

迭代次数	σ^2	τ_{00}	τ_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	1.008	2.193	0.451	1.105	1.239	1.006	0.997
50	1.008	0.565	0.450	1.778	2.785	1.033	0.984
500	1.008	0.553	0.450	1.842	2.932	1.031	0.980
1000	1.008	0.553	0.450	1.842	2.932	1.031	0.980

通过模拟计算可以直观的看出依据迭代次数 5，50，500，1000 的变化，分层线性回归模型 Guass-Seidel 型迭代算法所得到的参数估计值，可以看出在迭代 50 次后参数有明显的收敛趋势，在迭代 500 次后参数的估计值收敛到固定值，对应的参数估计值为$\widehat{\sigma}^2=1.008$，协方差估计值对应为$[\widehat{\tau}_{00},\widehat{\tau}_{01},\widehat{\tau}_{10},\widehat{\tau}_{11}]=[0.553,-0.040,-0.040,0.450]$，固定效应的估计值对应为$[\widehat{\gamma}_{00},\widehat{\gamma}_{01},\widehat{\gamma}_{10},\widehat{\gamma}_{11}]=[1.842,2.932,1.031,0.980]^\top$。

从表 5.1 和表 5.2 中可以看出，随着迭代次数的增加，分层线性模型 EM 算法和分层线性模型 Guass-Seidel 型迭代算法逐渐收敛，都表现出逐渐精确的趋势，可以在迭代 50 次时看出分层线性模型 Guass-Seidel 型迭代算法相比于分层线性模型 EM 算法中大部分的参数估计值有明显的收敛效果。

同时依据给出的收敛停止的判断准则，来对比分层线性模型 Guass-Seidel 型迭代算法和分层线性模型 EM 算法的收敛停止后的次数，来体现收敛速度。这里设置容忍度为 0.00005，设置迭代收敛的判断准则条件为第$i+1$次迭代值$\widehat{\sigma}^{2^{(i+1)}}$与第$i$次迭代值$\widehat{\sigma}^{2^{(i)}}$的差值小于容忍度，以及矩阵$\gamma$在迭代次数第$i+1$与第$i$次上对应所有元素的差值小于容忍度，以及协方差矩阵$T$在迭代次数第$i+1$与第$i$次上对应的所有元素的差值小于容忍度，这三个条件同时满足。同时由表 5.1 和表 5.2 体现出，这里设置的迭代次数上限为 500 次，具体得到的结果为：分层线性回归模型 EM 算法在第 218 次迭代停止，参数估计值依据判断准则收敛，具体得到的

收敛估计值为：$\widehat{\sigma}^2 = 1.007$，$[\widehat{\tau}_{00}, \widehat{\tau}_{01}, \widehat{\tau}_{10}, \widehat{\tau}_{11}] = [0.539, 0.083, 0.083, 0.595]$，$[\widehat{\gamma}_{00}, \widehat{\gamma}_{01}, \widehat{\gamma}_{10}, \widehat{\gamma}_{11}] = [1.950, 3.111, 1.067, 1.095]^{\top}$。分层线性回归模型 Guass-Seidel 型迭代算法在第 144 次迭代停止，参数估计值依据判断准则收敛，具体得到的收敛估计值为：$\widehat{\sigma}^2 = 0.930$，$[\widehat{\tau}_{00}, \widehat{\tau}_{01}, \widehat{\tau}_{10}, \widehat{\tau}_{11}] = [0.572, -0.143, -0.142, 0.455]$，$[\widehat{\gamma}_{00}, \widehat{\gamma}_{01}, \widehat{\gamma}_{10}, \widehat{\gamma}_{11}] = [2.058, 2.970, 0.920, 0.948]^{\top}$，由于设置的容忍度和随机数产生的原因，所以造成的参数估计值与上述表中的参数估计值有区别，不影响文章结论的判断结果。

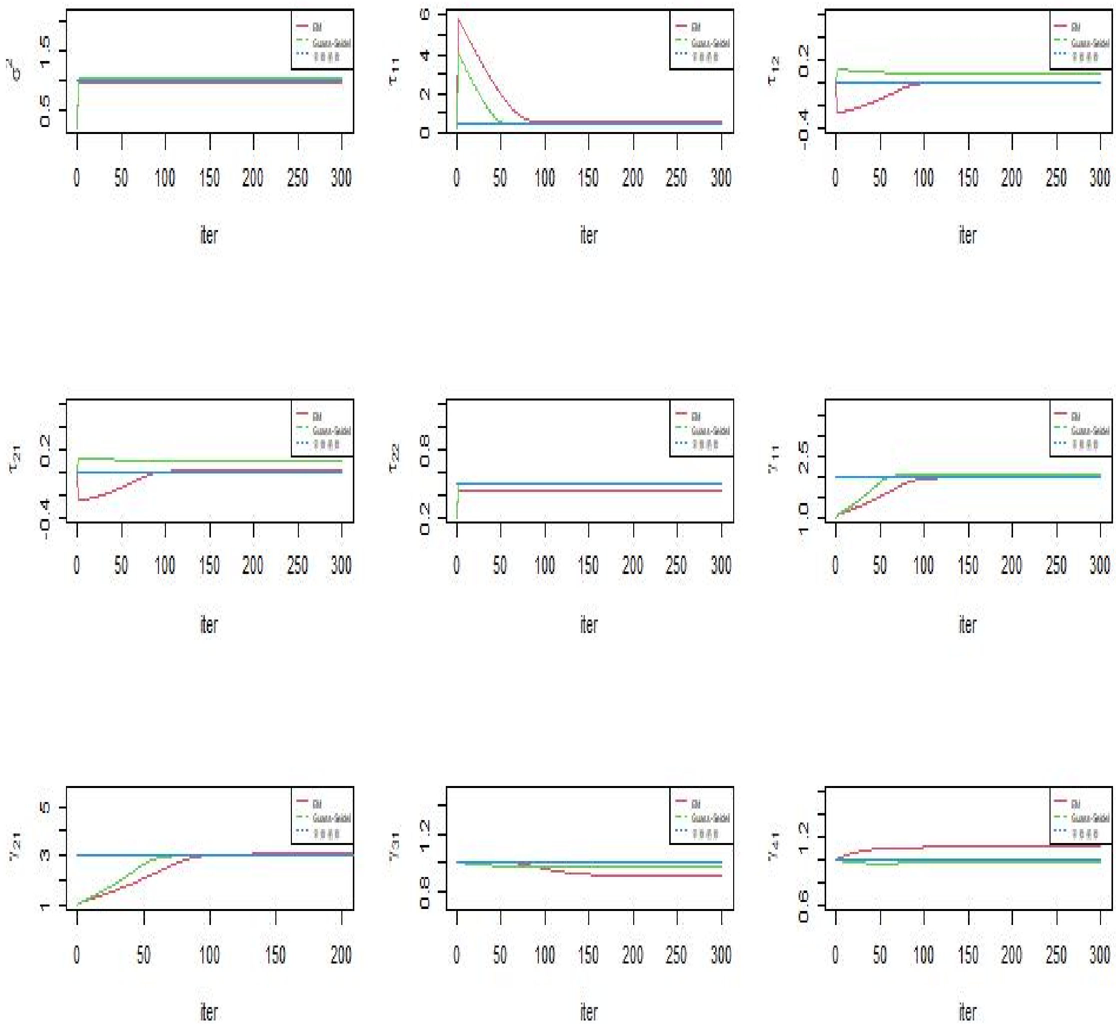


图 5.1 分层线性模型下 EM 算法与 Guass-Seidel 型算法迭代对比图

如图 5. 1，红色线对应为分层线性模型 EM 算法迭代过程，绿色线为分层线性模型 Guass-Seidel 型迭代算法过程，蓝色线表示参数真实值对应的常数函数，$iter$表示算法迭代次数。迭代次数依据上述迭代收敛的原理，这里设置包含初始值在内的 301 次迭代过程，从图中可以看出不论是分层线性 EM 算法还是分层线性 Guass-Seidel 型迭代算法在迭代过程中都存在收敛的趋势。同时可以看出分层线性模型 Guass-Seidel 型迭代算法在大多数参数上都比分层线性模型 EM 算法要提前收敛，即收敛速度较快，例如图形中对应参数σ^2，参数τ_{11}，参数τ_{12}，参数τ_{21}，参数τ_{22}，参数γ_{11}，参数γ_{21}等，都表现出参数收敛速度明显的特征，体现了分层模型 Guass-Seidel 型迭代算法相比于分层线性模型 EM 算法具有更好的收敛特征。

5. 2 案例分析

5. 2. 1 高校数学成绩数据

美国教育部的国家教育统计中心作为纵向研究项目的一部分，记录“随着他们开始承担成人的角色和责任，年轻人的教育、职业和个人发展”，数据来自于 160 所学校 7185 名学生数学成绩，采用分层线性回归模型对数据进行分析和解释，这里我们选取其中的部分数据进行嵌套分层线性回归模型分析。

对于分层数据中第一层水平即学生层面数据，含有 4 个变量如下：

MATHACH：数学成就的衡量标准，一般作为因变量，即Y_{ij}；

FEMALE：学生性别的指标(1 表示女性，0 表示男性)；

SES：学生的社会地位，由学生父母受教育程度、职业和收入等变量构成，变量已被标准化；

MINORITY：学生种族的指标(1 表示少数民族，0 表示其他)。

对于分层数据中第二层水平即学校层面数据，含有 6 个变量如下：

SIZE：学校招生人数；

MEANSES：包含在第一层水平数据中，每个学校学生的平均社会地位；

DISCLIM：衡量学科氛围的标准；

SECTOR：学校类型(1 表示天主教教会学校，0 表示公立学校)；

PRACAD：从事学术研究的学生的比例；

HIMNTY：学校招生少数民族学生比例描述（1 表示超过 40%少数民族学生，0 表示其他）；

5.2.2 数据分析

这里，选取部分数据作为嵌套模型来进行数据分析，其中构建限制性分层线性模型为：MATHACH 作为第j所学校第i个学生因变量Y_{ij}，SES 作为第一层水平下第j所学校第i个学生自变量X_{1i}，MEANSES 和 DISCLIM 作为第二层水平下第j所学校的自变量，即w_0和w_1，非限制性分层线性模型构建为 MATHACH 作为第j所学校第i个学生因变量Y_{ij}，SES 和 MINORITY 作为第一层水平下第j所学校第i个学生自变量X_{1i}和X_{2i}，MEANSES 和 DISCLIM 和 PRACAD 作为第二层水平下第j所学校的自变量，即w_0，w_1和w_2，根据上述变量构造具体模型形式如下：

原假设为：

第一层：$Y_{ij} = \beta_{0j} + \beta_{1j} * SES_{ij} + \varepsilon_{ij}$

第二层：$\beta_{0j} = \gamma_{00} + \gamma_{01} * MEANSES_j + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11} * DISCLIM_j$

将第二层模型带入第一层模型得到的混合效应模型为：

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01} * MEANSES_j + \gamma_{10} * SES_{ij} \\ + \gamma_{11} * DISCLIM_j * SES_{ij} + u_{0j} + \varepsilon_{ij} \end{aligned}$$

表 5.3 原假设分层模型中固定效应估计

固定效应		系数	标准误	$t-ratio$	$d.f.$	p值
β_0	γ_{00}	12.647	0.148	85.214	158	<0.001
	γ_{01}	5.866	0.359	16.322	158	<0.001
β_1	γ_{10}	2.206	0.108	20.333	7023	<0.001
	γ_{11}	0.604	0.117	5.182	7023	<0.001

表 5.4 原假设分层模型中随机效应估计

随机效应	标准误	方差成分	$d.f.$	χ^2	p值
u_{0j}	1.628	2.650	158	672.809	<0.001
ε_{ij}	6.072	36.873			

根据表 5.3 和表 5.4，可知原假设分层线性模型中估计的参数个数为 6，同时模型经过 6 次迭代对数似然函数变化值达到最小，对数似然值为-23267.58。结果给出了固定效应变量的系数估计值和随机效应的估计，相应的系数向量为 $[12.647, 5.866, 2.206, 0.604]$，$p$值可以看出估计值都通过了显著性检验。根据表 5.4 可知随机效应的参数估计结果，即第一层随机效应中的方差为 36.873。

备择假设为：

第一层：$Y_{ij} = \beta_{0j} + \beta_{1j} * MINORITY_{ij} + \beta_{2j} * SES_{ij} + \varepsilon_{ij}$

第二层：$\beta_{0j} = \gamma_{00} + \gamma_{01} * MEANSES_j + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11} * PRACAD_j$

$\beta_{2j} = \gamma_{20} + \gamma_{21} * DISCLIM_j$

将第二层模型带入到第一层中得到备择假设的混合效应模型：

$$\begin{aligned} Y_{ij} =& \gamma_{00} + \gamma_{01} * MEANSES_j + \gamma_{10} * MINORITY_{ij} \\ &+ \gamma_{11} * PRACAD_j * MINORITY_{ij} + \gamma_{20} * SES_{ij} \\ &+ \gamma_{21} * DISCLIM_j * SES_{ij} + u_{0j} + \varepsilon_{ij} \end{aligned}$$

表 5.5 备择假设分层模型中固定效应估计

固定效应		系数	标准误	$t-ratio$	$d.f.$	p值
β_0	γ_{00}	12.647	0.148	85.188	158	<0.001
	γ_{01}	5.866	0.359	16.321	158	<0.001
β_1	γ_{10}	-2.902	0.221	-13.116	7021	<0.001
	γ_{11}	2.348	0.833	2.818	7021	0.005
β_2	γ_{20}	1.954	0.109	17.953	7021	<0.001
	γ_{21}	0.480	0.116	4.137	7021	<0.001

表 5.6　备择假设分层模型中随机效应估计

随机效应	标准误	方差成分	$d.f.$	χ^2	p值
u_0	1.635	2.673	158	689.521	<0.001
ε_{ij}	5.998	36.981			

表 5.5 和表 5.6 中的结果可知备择假设分层线性模型中估计的参数个数为 8。同时经过 6 次迭代似然函数变化值达到最小，对数似然值为-23181.56。从表中可以看出对备择假设情况下分层线性回归模型中固定效应相对应的系数估计值，在显著性水平为 0.05 下都通过了显著性检验，固定效应是通过最小二乘估计得到的，对应为$[12.647, 5.866, -2.902, 2.348, 1.954, 0.480]$。从表 5.6 可知备择假设情况下分层线性模型随机效应的估计，可知第一层随机效应中对应的方差为 35.981。

表 5.7 原假设与备择假设情况下的拟合差异统计量

假设情况	拟合差异统计量	估计参数个数
原假设	46535.166	6
备择假设	46363.125	8

根据表 5.7 结果可得知备择假设情况下的拟合差异统计量相比原假设情况下的拟合差异统计量减少了 172.041，同时参数的差异个数为 2，对应自由度为 2 的卡方分布，172.041 达到了极其显著的水平。

通过计算可以得到嵌套模型分层线性模型的似然比统计量$\Lambda(x_1,...,x_n)$的值为 172，对应在显著性水平$\alpha = 0.05$和自由度为 2 的卡方分布下，根据给定拒绝域，拒绝原假设，即接受非限制性分层线性模型，说明从学校层面上引入从事学术研究的学生比例变量，从学生层面引入学生种族这一变量，对高校学生数学成绩有显著影响。

6 总结与展望

6.1 主要结论

本文针对分层线性模型的基本理论，研究的主要内容包括三部分，第一部分是基于分层线性模型 EM 算法的参数估计过程，并结合 Guass-Seidel 型迭代算法利用迭代最新值的思想，给出了分层线性回归模型 Guass-Seidel 型迭代算法的参数估计过程；第二部分是在分层半参数模型上给出了 Guass-Seidel 型迭代算法的推导内容，体现 Guass-Seidel 迭代算法的普遍性；第三部分是针对分层线性模型的统计推断问题，在具有嵌套结构的分层线性模型第一层系数诊断统计方法即为传统多元线性回归模型的统计推断问题基础上，给出具有嵌套结构的分层线性模型第二层系数诊断的统计推断方法，主要构造分层线性模型似然比统计量，对于引入变量是否保留的问题上给出理论方法。根据上述三部分主要内容得出，第一，相比于分层线性模型 EM 算法，分层线性模型 Guass-Seidel 型迭代算法具有更好的收敛效果；第二，通过在分层半参数模型上的推广，体现了 Guass-Seidel 型迭代算法在分层模型上的普遍性；第三，通过似然比统计量的检验，给出了嵌套分层线性模型的理论判断。最后，通过随机模型得出 Guass-Seidel 型迭代算法参数估计值具有更好的收敛性，并通过实例结果体现了具有嵌套结构的分层线性模型统计推断的可行性。

6.2 有待进一步探讨的问题

本文在研究过程中发现，使用分层线性模型 Guass-Seidel 型迭代算法解决 EM 算法收敛速度缓慢的问题时，对于初始值的设定并没有进行合理的安排来提高迭代算法的收敛速度，同时对扩展的分层半参数模型 Guass-Seidel 型迭代算法没有进行数值模拟，在研究迭代算法和统计推断过程中，都是基于同方差的情况下，有兴趣的同学可以尝试在异方差的情况下进行推广。此外，在实证分析中，可以尝试更多样本量进行分析，实现方法的更一般性和实用性。

参考文献

[1] Goldstein H.Multilevel Statistical Models (2 nd Ed.)[M].New York:John Wiley 1995.

[2] Laird N M.Random effects model for longitudinal data[J].Biometrics,1982, 38(4):963-974.

[3] Dempster A P,Tsutakawa D B R K.Estimation in Covariance Components Models[J].Journal of the American Statistical Association,1981,76(374):341-353.

[4] Little R J A,Rubin D B.Statistical Analysis with Missing Data [M].New York:Wiley and Sons,Inc.1987.

[5] Dempster A P,Laird N M and Rubin D B.Maximum Likelihood estimation from incomplete data via the EM algorithm [J].Journal of the Royal Statistical Society. Series B,1997,39:1-38.

[6] Wu C.On the Convergence Properties of the EM Algorithm[J].Annals of Statistics, 1983,11(1):95-103.

[7] Stephen W,Raudenbush S W,Bryk A S.Hierarchical Linear Models:Applications and Data Analysis Methods Second Edition[M].Journal of the American Statal Association,1997,88(421).

[8] Meng X L,Rubin D B.Maximum likelihood estimation via the ECM algorithm:A general framework[J].Biometrika,1993,80(2):267-278.

[9] Louis T A.Finding the Observed Information Matrix When Using the EM Algorithm[J].Journal of the Royal Statistical Society:Series B (Methodological), 1982,44:226-233.

[10] Vermunt J K.An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models[J].Statistica Neerlandica,2010, 58(2):220-233.

[11] Walker S.An EM algorithm for nonlinear random effects models[J].Biometrics, 1996,52(3):934-944.

[12] Chown J,UU Müller.Corrigendum:Detecting heteroscedasticity in non-arametric regression using weighted empirical processes[J].Journal of the Royal Statistical

Society Series B (Statistical Methodology),2019,81(4).

[13] Alfandari L,Hassanzadeh A,Ljubi I.An Exact Method for Assortment Optimization under the Nested Logit Model[J].European Journal of Operational Research,2021,291(4).

[14] Li C S,Lee S M,Yeh M S.A test for lack-of-fit of zero-inflated negative binomial models[J].Journal of statistical computation and simulation,2019,89(7-9): 1301 - 1321.

[15] 田茂再.高等分层分位回归建模理论[M].科学出版社,2015.

[16] 尚月强.一种网上求解线性方程组的 Guass-Seidel 并行迭代算法[J].贵州师范大学学报(自然版),2006(01):80-84.

[17] 乔舰,景平.多层线性模型的参数估计[J].统计与决策,2012(05):16-19.

[18] 卢玉桂.EM 算法在多层线性模型参数估计中的应用[D].广西大学,2013.

[19] 贺建风,付永超,熊健.基于分层贝叶斯广义线性模型的小域估计方法研究[J].数理统计与管理,2019(2):247-260.

[20] 赵必华,周元宽.大学生学校认同影响因素的多层线性模型分析[J].复旦教育论坛,2019(4):56-63.

[21] 王芳.基于分层线性模型的大学生教学满意度影响因素分析[J].复旦教育论坛,2018,16(01):48-55+97.

[22] 吴喜之.应用回归及分类:基于 R[M].中国人民大学出版社, 2016.

[23] 谢宇.回归分析.第 2 版[M].社会科学文献出版社, 2013.

[24] 田茂再, 陈歌迈. 条件分位中的分层线性回归模型[J].中国科学(A 辑:数学), 2006,10:25-40.

[25] 吴密霞.线性混合效应模型引论[M].科学出版社,2013.

[26] 刘红云.追踪数据分析方法及其应用[M].教育科学出版社,2005.

[27] 马海强.变系数模型的统计诊断和影响分析[D].中南大学,2008.

[28] 戴林送,林金官.广义泊松回归模型的统计诊断[J].统计与决策, 2013(21):29-33.

[29] 梁晋雯,田茂再.大数据下基于体积抽样的异常点诊断及估计问题[J].数理统计与管理,2020(2).

[30] 茆诗松,王静龙,濮晓龙.高等数理统计.第 2 版[M].高等教育出版社, 2006.

[31] Smith D V L F M .Bayes Estimates for the Linear Model[J].Journal of the Royal

Statistical Society,1972,34(1):1-41.

[32]Stadtmuller M U.Estimation of Heteroscedasticity in Regression Analysis[J]. Annals of Statistics,1987,15(2):610-625.

[33]Smith A F M.A general Bayesian Linear model[J].Journal of Royal Statistical Society,Series B,1973,35(1):67-75.

[34]Zhu H,Ibrahim J G,Shi X.Diagnostic Measures for Generalized Linear Models with Missing Covariates[J].Scandinavian Journal of Stats Theory & Applications, 2010,36(4):686-712.

[35]Dempster A P,Tsutakawa D B R K.Estimation in Covariance Components Models[J].Journal of the American Statistical Association,1981,76(374):341-353.

[36]卡塞拉,贝耶,张忠占,傅莺莺.统计推断:翻译版[M].机械工业出版社,2010.

[37]曹诗若,苏宇楠,田茂再.基于分层线性模型的贝叶斯推断及其应用[J].统计与决策,2015.

[38]姚远.基于分层线性模型的中高考数据挖掘[D].北京邮电大学,2019.

[39]晏振,戴晓文,田茂再.基于杠杆值大数据集抽样的异常点诊断[J].数理统计与管理, 2016, 35(5):794-802.

[40]曾婕,胡国治.Logistic 回归模型的统计诊断[J].数理统计与管理,2017, 36(004):620-631.

[41]孙广山.线性回归模型影响分析及异常点的统计诊断[D].东北林业大学, 2011.

[42]吕光明.线性回归模型中异常点的统计诊断[J].统计科学与实践, 2001,(06):22-23.

[43]费宇.线性和广义线性混合模型及其统计诊断[M].科学出版社, 2013.

[44]李柏椿.EM 算法及其改进算法在参数估计中的应用研究[D].重庆大学,2017.

致谢

桃李不言，下自成蹊，首先我要感谢我的指导老师田茂再老师，导师渊博的专业知识和严谨的治学态度，诲人不倦的高尚师德，严于律己和宽以待人的崇高风范，朴实无华、平易近人的人格魅力对本人影响深远。本论文从选题到完成，每一步都是在导师的悉心指导下完成的，在此，谨向导师表示崇高的敬意和衷心的感谢！

在研究生期间，同门师兄师姐师弟师妹们在导师的带领下体验了西北的迷人景色，师门不仅能在黄河边体验到美丽景色而且在路上的谈心与交流更让人幸福，老师不仅教会我们很多人生道理，而且在我们迷失方向时指引学习的道路，在丧失信心时给与支持和鼓励。在研究生期间，老师及同门带来的帮助不仅是学习上还是生活上，在阅读文献遇见问题时师姐会熬夜解答疑惑，遇见各种事情师姐都会支持我并给与我鼓励，师门探讨后会解决论文中出现的各种问题，线上讨论班更是体会到师门的智慧，以及要奋力追赶的勤奋和才华。在研究生的三年中，能遇到老师以及师兄师姐师弟师妹感到非常的荣幸，这段记忆将永远刻在脑海里。同时也非常荣幸遇到兰州财经大学的老师，老师们对学术的追求是我们永远学习的目标。

同时感谢我的室友们，给我的研究生生活带来了很多乐趣，是一大早互相督促去图书馆的动力，是中午一起恰饭的快乐，是夜晚在被窝里的八卦与谈心，是我遇到坎坷时给与心灵上的温暖，非常感谢我的室友们能容忍我的迷糊并教我成长。

研究生生活马上就要结束了，这意味着更多的未知和锻炼，我将带着老师和师门给予我的勇气，朋友给予的鼓励，继续奋斗前进，不断拼搏，创造人生的精彩。