微生物差异丰度分析方法的研究进展*

福建医科大学公共卫生学院流行病与卫生统计学系(350122)
林志锋　林少炜　饶雯清　付　蓉　林　征　胡志坚△

【中图分类号】 R181.2　　【文献标识码】 A　　DOI 10.11783/j.issn.1002-3674.2024.06.032

近年来,随着高通量基因组测序技术的发展,例如16S rRNA 基因靶向扩增子测序和 Shotgun 宏基因组测序,人们能够更加精确地定量分析人体内微生物组的组成信息[1]。重要的统计学任务之一是差异丰度(differential abundance, DA)分析,该分析基于测序处理获得的微生物丰度表与样本的元数据特征表,旨在确定与感兴趣的变量相关的微生物。DA 分析可以为疾病机制研究提供生物学见解,并探索可能作为疾病预防、诊断和治疗的生物标志物[2]。由于微生物丰度数据的复杂多样性,目前尚未开发出成熟稳健的 DA 分析工具,以确定可靠的微生物标志物。本文旨在对近年来国内外研究者广泛应用的 DA 分析工具进行整理,系统地总结各个方法的核心思想和优缺点,以便同领域研究者更方便、灵活地选择 DA 分析工具,并探讨新方法的发展方向。

微生物数据的特点

1.成分数据

微生物组数据是成分数据,即每个样本中的所有微生物相对丰度之和为 1[3]。虽然理想情况下,我们希望测量微生物的绝对丰度,即单位面积/体积的微生物数量,并对绝对丰度数据进行 DA 分析,但由于测序过程中涉及 DNA 的提取、PCR 扩增等复杂的化学过程,容易产生实验误差,较难获得标本中真实的总微生物数量[4]。尽管有几种实验技术(如 qPCR、spike-in 和流式细胞术等)可以获得微生物绝对丰度,但这些技术仍存在各自的局限性[5],目前主要的测序方法仍然只能提供相对丰度信息。然而,基于相对丰度数据推断未知绝对丰度变化是具有挑战性的,因为某些细菌分类群的相对丰度相对于感兴趣的协变量的增加或减少会自动导致其他细菌分类群的相对丰度变化,这种统计学现象叫成分效应。当在大量低丰度分类群中存在几个高丰度分类群时,微生物组的成分效应更加明显[6]。

2.零膨胀

微生物组丰度数据通常存在大量的零计数。在典型的微生物组数据集中,超过 70% 的微生物丰度为零,这种情况被称为零膨胀。零膨胀的原因可能是物理缺失(结构零)或采样不足(采样零)[7]。结构零是由于某些微生物分类群在大多数样品中不存在而无法检测到,而采样零则是因为低丰度微生物的存在与否高度依赖于测序深度。此外,DNA 偏倚、批量效应或聚合酶链反应偏倚等实验原因也容易导致采样零点。

3.稀疏性

16S rRNA 基因扩增子测序获得的序列通常根据相似性(如 97%)聚类为操作分类单元(operational taxonomic unit, OTU),并将其用于下游统计分析。同一 OTU 在不同样本中的计数值可能不同,由于微生物群落中主导菌的存在,使得少数 OTU 在大多数样本中出现频率较高,而其余大部分的 OTU 都比较稀疏,一些高度稀疏的 OTU 计数值接近于 0,只能检测到很小的比例[8]。

4.过度分散

微生物组测序实验中的 reads 数或 OTU 计数通常会呈现过度分散的情况。例如在单个样本中,某些 OTU 计数值很小,甚至趋近于 0,而某些 OTU 计数值很大,甚至达到几千,总体数据呈现严重的右偏态分布。读取计数的均值大于泊松分布预测的方差,即存在偏大离差的现象。

5.高维度

微生物组丰度数据是高维数据集,通常包含数百至数千个不同的 OTU[9]。由于微生物类别众多且它们所包含的数量参差不齐,即使在仅有几十个样本的情况下,其测序得出的微生物“属”的种类也可以达到上百种。这种情况下,样本数量通常小于或远远小于微生物种类的数量。

微生物数据差异丰度分析方法

近年来,微生物数据 DA 分析方法涌现出越来越多的创新和发展。其中,主要围绕成分数据分析、零膨胀分析、线性模型等方面进行方法的优化创新和开发,表 1 列出了目前部分主要的 DA 分析方法。

* 基金项目:福建省科技创新联合资金(2020Y9018);福建省自然科学基金(2021J01726)
△通信作者:胡志坚,E-mail:huzhijian@ fjmu.edu.cn

表 1 主要的 DA 分析方法汇总

成分数据分析方法			零膨胀数据分析方法	纵向数据分析方法	其他方法		
对数比转换	归一化	内参法			早期经典	线性模型	中介分析
ANCOM	edgeR	DACOMP	ZIP	ZINB	LEfSe	LDM	MarZIC
ANCOM-Ⅱ	DESeq2	RAIDA	ZIB	ZIBR	STAMP	LOCOM	LDM-med
fastANCOM	metagenomeSeq	RioNorm2	ZINB	ZIGMM		LinDA	PERMANOVA-med
adaANCOM	Omnibus test		ZIGLM	FZINBMM			
ANCOM-BC			ZIBB				
ALDEx			NBZIMM				
ALDEx2							

1.成分数据分析方法

早在 1896 年,Pearson 指出,如果用传统的方法计算比例数据的相关性,会出现伪相关[10],这是由于成分数据具有定和约束,数据存在于单纯形空间而非欧几里得空间上,因此相应的统计分析可能存在困难。如果使用经典的统计方法,如两样本 *t* 检验、Wilcoxon 秩和检验、线性回归分析等,忽视数据的成分性质,则会导致大量的错误结果。

微生物成分数据分析方法的开发需要解除成分限制,通常是将读取计数数据的非零部分进行某种形式的对数比转换,如加性对数比(additive log-ratio, alr)转换或中心对数比(centered log-ratio, clr)转换[11]。而不能进行对数变换的零计数数据,通常的做法是在分类群计数表的零或所有条目中添加一个伪计数,最常见的是 1 或 0.5 或更小的数值[3,12-13]。然而,选择伪计数的方法并不统一,而且已经证明伪计数转换会影响成分分析的结论[14]。

常见的微生物组成分析方法包括微生物成分分析(analysis of composition of microbiomes, ANCOM)[12]、ANCOM-Ⅱ[15]、fastANCOM[16]、adaANCOM[17]、ANCOM-BC[13]、类方差差异表达(ANOVA-like differential expression, ALDEx)[18]、ALDEx2[19]、DACOMP[20]等。ANCOM 是一种基于 alr 转换的方法,通过逐个使用每个分类单元作为参考分类单元来执行所有可能的 DA 分析,当 *t* 检验和零膨胀分析[21]的错误发现率(false discovery rate, FDR)在 60%以上时,ANCOM 总能将 FDR 控制在 5%水平以下。但由于计算时间与分类单元的数量成正比,所以计算密集且运行时间较长。ANCOM-Ⅱ是 ANCOM 的升级,考虑了采样零与结构零的问题。针对 ANCOM 运行时间久的缺点,Zhou 等人[16]改进了模型算法,提出了 fastANCOM,计算速度提升了约近百倍。为了解决高维环境下微生物数据计算的复杂性,Zhou 等人[17]提出了自适应微生物组成分析方法(adaANCOM),它首先将 Dirichlet-Tree 多项式(Dirichlet-tree multinomial, DTM)扩展为零膨胀 DTM,用于微生物数据的多元建模。其次,在此框架内使用贝叶斯公式,引入后验均值变换将原始计数转换为和为 1 的非零相对丰度。接着,利用变换后的数据,在分类树上自适应地构造对数比来进行 DA 分析。adaANCOM 巧妙地解决了微生物数据零膨胀、高维度、成分数据的特点,计算效率较高,同时又较好地控制了 FDR。ANCOM-BC 使用基于偏差的对数线性模型对观测丰度进行建模,明确地检验了关于单个分类单元差异绝对丰度的假设,同时估计了样本特定的抽样分数,并适当地纠正由于成分效应导致的不平等采样所造成的偏差。

ALDEx 和 ALDEx2 是基于类方差分析思想的工具,他们的核心思想是将样本间的变异分解为四个部分:条件内变异、条件间变异、抽样变异和一般误差。通过对这些变异的分析,ALDEx 和 ALDEx2 可以用来识别两个或多个组之间的特征差异丰度。在 ALDEx2 中,分类单元丰度的后验概率被融入到一个组成框架中,通过 Dirichlet 分布和蒙特卡罗采样过程,将观察到的丰度转换为相对丰度,后对相对丰度矢量进行 clr 转换,再最后进行显著性检验和 *P* 值校正,识别差异丰度的特征。ALDEx2 具有较稳健的 FDR 控制水平和较好的检验功效[22]。

DACOMP[20]、RAIDA[23]和 RioNorm2[24]利用了参考分类群的方法进行数据分析,这些方法的基本思想是找到相对于感兴趣的变量来说不变化的一个或一组参考分类群,然后利用 taxa/参考分类群的丰度比来进行 DA 分析。DACOMP 选择一组在 DA 分析前最不可能出现差异的参考类群[25]。RAIDA[23]找到一个在 DA 分析中发现最少的参考类群。而 RioNorm2[24]依靠基于网络的归一化来寻找相对不变的分类群。这些基于内参比的方法可以有效地解除成分数据的限制,但该类方法的挑战在于确定理想的参考分类群较为困难。

此外,一些归一化方法已经被广泛应用于微生物 DA 分析,例如 edgeR[26]、DESeq2[27]、metagenomeSeq[28]和 Omnibus test[29],这些方法使用了不同的归一化技术,分别包括 *M* 值的修饰平均数(trimmed mean of M-values, TMM)、相对对数表达(relative log expression, RLE)、累积总和缩放(cumulative sum scal-

ing，CSS）和对数比的几何平均值（geometric mean of pairwise ratios，GMPR）[30]，这些方法都证明可以有效地应用于微生物组成数据的分析。

2.零膨胀分析方法

微生物丰度数据中，出现 0 的个数要明显多于泊松、二项或负二项等标准离散分布随机产生的个数。在过去的十年中，零膨胀模型已经成为分析零过多数据的有效方法，例如零膨胀泊松模型（zero-inflated poission，ZIP）、零膨胀二项模型（zero-inflated binomial，ZIB）、零膨胀负二项模型（zero-inflated negative binomial，ZINB）和零膨胀广义线性模型（zero-inflated generalized linear model，ZIGLM）等[31]。这些经典模型也广泛应用于微生物数据 DA 分析。

近年来，越来越多的研究者基于经典的零膨胀模型进行改进优化，以更好地适应具有复杂特点的微生物数据。Fang 等人[32]将 ZINB 方法应用于 DA 分析，用于识别两个或多个微生物群落之间的差异类群。该模型由两部分组成：二项分布解释过度分散，零膨胀模型解释多余的零。对存在零过多，过度离散的重复测量的计数数据，也可采用 ZINB 模型分析[33]。零膨胀的 β-二项式（zero-inflated β-binomial，ZIBB）[34]利用均值-方差关系来提高功效并调整协变量，该方法也是一个混合模型，包括解释过量零的模型和一个计数模型，并通过 β-二项式回归解释过度分散。负二项零膨胀混合模型（negative binomial and zero-inflated mixed model，NBZIMM）[35]提供了设置和拟合负二项式混合模型、零膨胀负二项式混合模型和零膨胀高斯混合模型的功能，NBZIMM 解决了微生物组研究中的数据特征和复杂设计的问题。近年来，零膨胀混合模型的开发备受关注，这些模型更好地解决了零膨胀、过度分散和稀疏性等问题。

3.基于纵向数据的分析方法

广义估计方程（generalized estimating equations，GEEs）和广义线性混合效应模型（generalized linear mixed model，GLMM）是分析纵向数据最主要的两种方法[36]。这些模型也应用于微生物组研究中，例如分析孕妇和非孕妇之间微生物组组成和稳定性的差异[37]。在此基础上，也有研究者结合微生物零膨胀、过度分散等特点，提出零膨胀相关的纵向模型，例如 ZINB、零膨胀 β 回归（zero-inflated β regression，ZIBR）[38]、零膨胀高斯混合模型（zero-inflated Gaussian mixture model，ZIGMM）[39]、快速零膨胀负二项混合模型（fast zero-inflated negative binomial mixture model，FZINBMM）[40]等。ZIBR 应用于相对丰度数据，通过伯努利分布捕捉微生物的存在或不存在，同时使用 β 分布捕捉非零丰度，它实际上可以被认为是 logistic 回归和 β 回归分量的混合，该模型可以评估每个分类单元的丰度随时间和组间变化的情况。ZIGMM 是 logistic 回归和高斯回归分量的混合而成的模型，它可以评估时间效应、群体效应和时间×群体相互作用效应。FZINBMM 用于计数数据，使用 EM-IWLS 算法进行拟合，不仅可以评估时间和群体效应，还考虑了计数数据过度分散和稀疏性的特点。通过模拟数据和实际数据的分析比较，FZINBMM 算法在性能上优于 ZIBR 和 ZIGMM 算法。

4.其他分析方法

LEfSe[41]和 STAMP[42]是早期应用于微生物数据 DA 分析的方法，但由于缺乏对微生物数据特点的考虑，假阳性率较高，不能很好地控制 FDR。因此，随着新方法的涌现，它们逐渐被替代。

近些年，基于线性模型的 DA 方法开发备受关注。我国学者吴晶晶等人[43]对于微生物组学中的高维成分数据分析方法进行了总结，较为详细介绍了线性对数比模型、结合树结构的高维线性回归模型和树结构惩罚函数模型等。李玉莹等人[44]提出了带有等式约束的成分数据的 logistic 回归模型，并将其应用于肠道菌群与 2 型糖尿病的研究中，数据模拟和实证表明该模型有良好的正确性、准确性和有效性。Hu 等人[45]引入了线性分解模型（linear decomposition model，LDM），该模型可以全局测试微生物组的任何影响，并对单个 OTU 的影响进行测试。模拟数据分析表明 LDM 具有良好的检验性能，并且可以很好地控制 FDR。此外，Hu 等人[46]还提出了一种名为 LOCOM 的稳健的 logistic 回归方法，该方法基于两个类群中比较计数的 log 优势比对实验偏差是不变的原理，不需要伪计数即可进行分析成分数据。Zhou 等人[47]提出 LinDA 用于微生物组组成数据的 DA 分析。该方法基于 clr 转换的丰度数据识别了与传统线性回归模型相关的偏差，并提出了估计和纠正偏差的策略。LinDA 可以推广到线性混合效应模型，用于更复杂的数据和研究设计，同时也适用于其他高维成分数据的 DA 分析。

Wu 等人[48]开发了边缘中介分析方法（MarZIC）来解释微生物组数据的组成结构，以零膨胀微生物组组成作为介质进行分析时，MarZIC 方法优于标准的因果中介分析方法。另外，Yue 等人[49]在原有的 LDM 模型上，发明了微生物中介分析方法：LDM-med 和 PERMANOVA-med，分别探讨了单个分类水平和群落水平上的中介分析问题。

现有 DA 分析方法的局限性及展望

本文综述了当前可用于微生物组数据分析的统计方法和模型。包括经典的统计方法和模型以及最近几年发展的新方法。新开发的方法主要针对微生物组数

据的特定特征:例如成分数据、零膨胀、过度分散、高维度等。然而,现有的 DA 工具仍有其局限性:①目前微生物成分数据分析仍未解决零值问题。现有主要分析策略是添加伪计数,但是添加伪计数是否会改变结果很难得到检验。②不考虑成分数据影响,当前的计数模型具有灵活的调整能力,主要针对微生物组的高维数据结构、零膨胀、过度分散和稀疏性等特点,更适合于微生物组数据的统计和生物学建模,但许多模型仍然需要提高联合建模的能力。③纵向研究可以捕捉微生物组内部随时间改变的变化情况,从而深入了解微生物系统。然而,由于数据的复杂性和固有特性,现有的统计方法受到了限制[50]。因此,仍需开发新的纵向模型以适应微生物组、环境和宿主之间的复杂时间相关性。④在检测因果关系和因果推断方面,中介分析仍处于初级阶段。为了满足对动态复杂的微生物组数据建模的需求,需要合适的统计工具来分析假设因素之间的因果关系和中介关系。

近年来,微生物组数据的统计分析取得了很大的进展,但是新的统计方法和模型仍有发展的空间。新统计方法的重点可以聚焦在以下几个方面:①考虑到微生物组数据的组成性质,并同时解决了高维度、零膨胀、过度分散和稀疏性等特点。②不考虑微生物数据组成性质,继续开发适当的计数模型,以共同拟合和有效地解释复杂的微生物组数据的特点。③继续开发纵向模型和因果模型,以便更准确地推断微生物群、环境和宿主之间的动态和复杂联系。

参 考 文 献

[1] 刘永鑫,秦媛,郭晓璇,等.微生物组数据分析方法与应用[J].遗传,2019,41(9):845-862.

[2] Zhou Y, Xu ZZ, He Y, et al. Gut Microbiota Offers Universal Biomarkers across Ethnicity in Inflammatory Bowel Disease Diagnosis and Infliximab Response Prediction [J]. mSystems, 2018, 3(1): e00188-17.

[3] Gloor GB, Macklaim JM, Pawlowsky-Glahn V, et al. Microbiome Datasets Are Compositional: And This Is Not Optional [J]. Frontiers in Microbiology, 2017, 8: 2224.

[4] Tsilimigras MCB, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges [J]. Annals of Epidemiology, 2016, 26(5): 330-335.

[5] Morton JT, Marotz C, Washburne A, et al. Establishing microbial composition measurement standards with reference frames [J]. Nature Communications, 2019, 10(1): 2719.

[6] Quinn TP, Erb I, Richardson MF, et al. Understanding sequencing data as compositions: an outlook and review [J]. Bioinformatics (Oxford, England), 2018, 34(16): 2870-2878.

[7] Silverman JD, Roche K, Mukherjee S, et al. Naught all zeros in sequence count data are the same [J]. Computational and Structural Biotechnology Journal, 2020, 18: 2789-2798.

[8] Calgaro M, Romualdi C, Waldron L, et al. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data [J]. Genome Biology, 2020, 21(1): 191.

[9] NIH HMP Working Group, Peterson J, Garges S, et al. The NIH Human Microbiome Project [J]. Genome Research, 2009, 19(12): 2317-2323.

[10] Pearson K. On a form of spurious correlation which may arise when indices are used in the measurement of organs [C]. London: Proc, 1897, 60: 489-502.

[11] Aitchison J. The Statistical Analysis of Compositional Data [J]. Journal of the Royal Statistical Society: Series B (Methodological), 1982, 44(2): 139-160.

[12] Mandal S, Van Treuren W, White RA, et al. Analysis of composition of microbiomes: a novel method for studying microbial composition [J]. Microbial Ecology in Health and Disease, 2015, 26: 27663.

[13] Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction [J]. Nature Communications, 2020, 11(1): 3514.

[14] Costea PI, Zeller G, Sunagawa S, et al. A fair comparison [J]. Nature Methods, 2014, 11(4): 359.

[15] Kaul A, Mandal S, Davidov O, et al. Analysis of Microbiome Data in the Presence of Excess Zeros [J]. Frontiers in Microbiology, 2017, 8: 2114.

[16] Zhou C, Wang H, Zhao H, et al. fastANCOM: a fast method for analysis of compositions of microbiomes [J]. Bioinformatics, 2022, 38(7): 2039-2041.

[17] Zhou C, Zhao H, Wang T. Transformation and differential abundance analysis of microbiome data incorporating phylogeny [J]. Bioinformatics, 2021, 37(24): 4652-4660.

[18] Fernandes AD, Macklaim JM, Linn TG, et al. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq [J]. PLoS One, 2013, 8(7): e67019.

[19] Fernandes AD, Reid JN, Macklaim JM, et al. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis [J]. Microbiome, 2014, 2(1): 15.

[20] Brill B, Amir A, Heller R. Testing for Differential Abundance in Compositional counts data, with Application to Microbiome Studies [J]. Annals of Applied Statistics, 16(4): 2648-2671.

[21] Paulson JN, Stine OC, Bravo HC, et al. Differential abundance analysis for microbial marker-gene surveys [J]. Nature Methods, 2013, 10(12): 1200-1202.

[22] Nearing JT, Douglas GM, Hayes MG, et al. Microbiome differential abundance methods produce different results across 38 datasets [J]. Nature Communications, 2022, 13(1): 342.

[23] Sohn MB, Du R, An L. A robust approach for identifying differentially abundant features in metagenomic samples [J]. Bioinformatics (Oxford, England), 2015, 31(14): 2269-2275.

[24] Ma Y, Luo Y, Jiang H. A novel normalization and differential abundance test framework for microbiome data [J]. Bioinformatics (Oxford, England), 2020, 36(13): 3959-3965.

[25] Brill B, Amir A, Heller R. Testing for differential abundance in compositional counts data, with application to microbiome studies [J]. Annals of Applied Statistics, 16(4): 2648-2671.

[26] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data [J]. Bioinformatics, 2010, 26(1): 139-140.

[27] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 [J]. Genome Biology, 2014, 15(12): 550.
[28] Kumar MS, Slud EV, Hehnly C, et al. Differential richness inference for 16S rRNA marker gene surveys [J]. Genome Biology, 2022, 23(1): 166.
[29] Chen J, King E, Deek R, et al. An omnibus test for differential distribution analysis of microbiome sequencing data [J]. Bioinformatics (Oxford, England), 2018, 34(4): 643-651.
[30] Chen L, Reeve J, Zhang L, et al. GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data [J]. Peer J, 2018, 6: e4600.
[31] 解锋昌,韦博成,林金官.ZI 数据的统计分析综述[J].应用概率统计,2009,25(6):659-671.
[32] Fang R, Wagner BD, Harris JK, et al. Zero-inflated negative binomial mixed model: an application to two microbial organisms important in oesophagitis [J]. Epidemiology and Infection, 2016, 144(11): 2447-2455.
[33] 赵丽华,刘桂芬,田娇妮.重复测量计数资料的随机效应 ZINB 模型[J].中国卫生统计,2011,28(6):665-667.
[34] Hu T, Gallins P, Zhou YH. A Zero-inflated Beta-binomial Model for Microbiome Data Analysis [J]. Stat(International Statistical Institute), 2018, 7(1): e185.
[35] Zhang X, Yi N. NBZIMM: negative binomial and zero-inflated mixed models, with application to microbiome/metagenomics data analysis [J]. BMC Bioinformatics, 2020, 21(1): 488.
[36] 汤宁,宋秋月,易东,等.医学纵向数据建模方法及其统计分析策略[J].中国卫生统计,2019,36(3):441-444+447.
[37] Romero R, Hassan SS, Gajer P, et al. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women [J]. Microbiome, 2014, 2(1): 4.
[38] Chen EZ, Li H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data [J]. Bioinformatics(Oxford, England), 2016, 32(17): 2611-2617.
[39] Zhang X, Guo B, Yi N. Zero-Inflated gaussian mixed models for analyzing longitudinal microbiome data [J]. PloS One, 2020, 15(11): e0242073.
[40] Zhang X, Yi N. Fast zero-inflated negative binomial mixed modeling approach for analyzing longitudinal metagenomics data [J]. Bioinformatics, 2020, 36(8): 2345-2351.
[41] Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation [J]. Genome Biology, 2011, 12(6): R60.
[42] Parks DH, Tyson GW, Hugenholtz P, et al. STAMP: statistical analysis of taxonomic and functional profiles [J]. Bioinformatics, 2014, 30(21): 3123-3124.
[43] 吴昌晶,何顺,邓明华.微生物组学中的高维计数和成分数据分析[J].中国科学:数学,2017,47(12):1735-1760.
[44] 李玉莹,张景肖.成分数据的 logistic 回归模型研究[J].数理统计与管理,2019,38(3):442-449.
[45] Hu YJ, Satten GA. Testing hypotheses about the microbiome using the linear decomposition model(LDM)[J]. Bioinformatics(Oxford, England), 2020, 36(14): 4106-4115.
[46] Hu Y, Satten GA, Hu YJ. LOCOM: A logistic regression model for testing differential abundance in compositional microbiome data with false discovery rate control [J]. Proceedings of the National Academy of Sciences, 2022, 119(30): e2122788119.
[47] Zhou H, He K, Chen J, et al. LinDA: linear models for differential abundance analysis of microbiome compositional data [J]. Genome Biology, 2022, 23(1): 95.
[48] Wu Q, O'Malley J, Datta S, et al. MarZIC: A Marginal Mediation Model for Zero-Inflated Compositional Mediators with Applications to Microbiome Data [J]. Genes, 2022, 13(6): 1049.
[49] Yue Y, Hu YJ. A New Approach to Testing Mediation of the Microbiome at Both the Community and Individual Taxon Levels [J]. Bioinformatics(Oxford, England), 2022, 38(12):3173-3180.
[50] Kodikara S, Ellul S, Lê Cao KA. Statistical challenges in longitudinal microbiome data analysis [J]. Briefings in Bioinformatics, 2022, 23(4): bbac273.

(责任编辑:郭海强)

(上接第 955 页)

[76] Lee S, Honavar V. Transportability of a Causal Effect from Multiple Environments[C].Proceedings of the Proceedings of the 27th Conference on Artificial Intelligence, 2013.
[77] Peters J, Bühlmann P, Meinshausen N. Causal inference using invariant prediction: identification and confidence intervals [J].Journal of the Royal Statistical Society Series B-Statistical Methodology, 2015, 78(5): 947-1012.
[78] Prosperi M, Guo Y, Sperrin M, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare [J]. Nature Machine Intelligence, 2020, 2(7): 1-7.
[79] Schwartz S, Gatto NM, Campbell UB. Transportability and Causal Generalization [J]. Epidemiology, 2011, 22(5): 745-746.
[80] García-Peña C, Ramírez-Aldana R, Parra-Rodriguez L, et al. Network analysis of frailty and aging: Empirical data from the Mexican Health and Aging Study [J]. Experimental Gerontology, 2019, 128: 110747.

(责任编辑:郭海强)