



3种成分数据方法在24 h活动行为时间使用流行病学研究中的应用比较

Comparison of Three Compositional Data Methods Applied to the 24 h Movement Time-Use Epidemiological Research

刘晴^{1,2}, 谭健怡², 黄宝莹², 任志强², 程金群², 黄珍惠²,
徐华富², 李红娟³, 郜艳晖^{4*}

LIU Qing^{1,2}, TAN Jianyi², HUANG Baoying², REN Zhiqiang², CHENG Jinqun²,
HUANG Zhenhui², XU Huaifu², LI Hongjuan³, GAO Yanhui^{4*}

摘要: 通过文献资料调研和实例分析比较等距对数比转换(isometric logratio transformation, ilr)、Dirichlet分布和Flexible Dirichlet (FD) 3种成分数据方法优缺点,旨在选择最优方法以阐明人群24 h活动行为时间使用比例分布。研究发现,相比于Dirichlet和FD方法,ilr更能准确描述研究人群24 h活动行为时间使用比例分布。未来的研究仍需继续探索成分数据方法,促进时间使用流行病学方法学研究及应用。

关键词: 身体活动;时间使用流行病学;成分数据;等距对数比转换;Dirichlet分布;柔性狄利克雷

Abstract: The study compared the advantages and disadvantages of three compositional data methods, including the isometric logratio transformation (ilr), Dirichlet distribution, Flexible Dirichlet (FD) and conducted case analysis. The aims were to choose the best method to clarify the distribution of 24 h movement time-use ratio in population. The study found that compared with Dirichlet and FD methods, ilr could more accurately describe 24 h movement time-use ratio distribution of population in this study. Future research explored compositional data methods was still necessary for promoting the research and application of time-use epidemiological methodology.

Keywords: physical activity; time-use epidemiology; compositional data; ilr; Dirichlet distribution; FD

中图分类号: G804.49 **文献标识码:** A

基金项目:

国家社会科学基金项目(ALA190015)

第一作者简介:

刘晴(1995-),女,研究实习员,硕士,主要研究方向为体质健康、流行病学统计方法,E-mail:liuqing_gd@163.com。

*通信作者简介:

郜艳晖(1973-),女,博士,教授,主要研究方向为流行病学统计方法,E-mail:gao_yanhui@163.com。

作者单位:

1. 广东省体育科学研究所,广东广州 510663;
2. 广东药科大学,广东广州 510310;
3. 北京体育大学,北京 100084;
4. 暨南大学,广东广州 510632
1. Guangdong Provincial Institute of Sports Science, Guangzhou 510663, China;
2. Guangdong Pharmaceutical University, Guangzhou 510310, China;
3. Beijing Sport University, Beijing 100084, China;
4. Jinan University, Guangzhou 510632, China.

近年来,身体活动(physical activity, PA)不足(卫文,2018;WHO,2018)、久坐行为(sedentary behavior, SB)增加(郭强等,2017;Bauman et al.,2011)和睡眠(sleep, SLP)不足(闫静弋等,2019;Keyes et al.,2015;Sheehan et al.,2019)等不良生活行为越来越普遍,导致肥胖、高血压、心血管疾病和癌症等风险升高,并增加全死因死亡率(于洪军等,2013;周志雄,2015;Carson et al.,2016;Chaput et al.,2016;Patterson et al.,2018;Poitras et al.,2016)。此外,过度的SB和SLP不足还会削弱身体活动带来的健康收益(Chaput et al.,2017)。既往研究多从绝对量角度探讨单个活动行为时间的分布特征或其与健康结局的关系,忽略了行为时间的相对大小及内在联系。从连续活动谱(movement continuum)角度来看,中高强度身体活动(moderate-to-vigorous-intensity physical activity, MVPA)、低强度身体活动(light-intensity physical activity, LPA)、SB和SLP构成一天24 h的活动行为,四者所用时间为互斥关系,即24 h内一种行为使用时间的改变,需在其他至少一种行为时间上进行补偿,所导致的健康效应除受该行为时间改变影响外,也部分归因于其他至少一种行为时间的补偿性改变。因此,将各种行为时间看作一个24 h连续谱,从相对量角度研究24 h各种行为时间的分布特征、影响因素以及与健康指标的关系

更趋合理,同时对制定 24 h 各种活动行为时间平衡指南,促使健康效益达到最优化也有重要意义。

目前,PA、SB 和 SLP 的 24 h 综合研究备受国内外学者关注。Pedišić(2014)认为,要确定 SLP、MVPA、LPA 以及总 PA 时间对健康的独立贡献,需要在这些变量间相互调整;作为时间使用成分的一部分,需要使用成分数据分析(compositional data analysis, CoDA)方法。宋俊辰等(2020)综述了时间使用流行病学在身体活动领域的应用进展。时间使用流行病学主要研究人群中与健康相关的时间使用模式的分布现状、趋势、决定因素及对健康结局的影响,以达到预防不良的时间使用模式,实现个体健康时间最佳分配的目标。这一概念由 Pedišić 等(2017)提出,同时提出的还有时间使用流行病学综合研究框架,时间使用的分布特征位于该研究框架中心位置,最佳时间使用平衡为核心要素,不同的结局可能与不同的时间使用模式有关。因此,准确描述 24 h MVPA-LPA-SB-SLP 使用时间的人群的分布特征,有助于确定最佳时间使用平衡模型,为制定公共卫生和健康干预策略提供科学可靠的参考和依据。

从统计学角度,24 h 活动行为研究除时间绝对量外,也显示了时间使用的相对量信息,对应的各活动行为时间数据存在定和限制(即所有成分比例总和为 1),称为成分数据(compositional data)。成分数据样本空间属于标准的单形空间,协方差阵至少有 D 个元素必为负数,其中 D 为成分个数(张尧庭,2000; Van den Boogaart et al., 2013)。因此,传统基于欧式空间的经典统计方法并不适用于成分数据。针对成分数据的统计方法最常用的是等距对数比变换(isometric logratio transformation, ilr),即假设成分数据服从加法逻辑正态分布(additive logistic normal, aln),通过 ilr 将其从单形空间映射到欧式空间,进而基于正态分布采用传统方法分析。现有研究采用 ilr 方法主要探讨 24 h 活动行为的各类健康效应,此时 24 h 活动行为被作为解释变量。但当需要描述 24 h 活动行为时间使用的分布,或将其作为结局变量时,ilr 方法易受异常值影响,可能会严重误导结果。更重要的是,ilr 方法得到的参数只能在转换后的空间下解释,没有直接的含义,结果解释不直观。相比于 ilr,Dirichlet 分布能直接基于成分数据单形空间定义模型,更直观、更有吸引力。该分布具有良好的数学性质和参数易解释等诸多优良性质(Ng et al., 2011),但对于成分间强相关或弱正相关的成分数据适用性较差。鉴于此, Ongaro 等(2013)提出 Dirichlet 的混合分布,即 Flexible Dirichlet(FD)模型,既能保留 Dirichlet 分布的良好数学和成分性质,又可弥补 Dirichlet 分布不足,具有更大的灵活性。相比于 ilr 和 Dirichlet 分布,FD 模型更能捕获成分数据的复杂特征(Migliorati et al., 2017a),但目前 Dirichlet 或 FD 方法在 24 h 使用流行病学研究中仍鲜有应用。

本研究详细介绍了 3 种成分数据方法及其优缺点,并通过实例分析比较 3 种方法在拟合 24 h 活动行为时间分布的效果,选择最优方法阐明 24 h PA、SB 和 SLP 行为时间使用分布特征。

1 成分数据原理与方法

1.1 成分数据原理

若任意 D 元向量 $\mathbf{y}=(y_1, y_2, \dots, y_D)$ 满足以下表达式:

$$\mathcal{S}^D = \left\{ \mathbf{y} = (y_1, y_2, \dots, y_D) : y_i \geq 0; i = 1, 2, \dots, D; \sum_{i=1}^D y_i = 1 \right\} \quad (1)$$

式(1)中, \mathbf{y} 为 D 元成分数据,记为 $\mathbf{y} \in \mathcal{S}^D$ 。D 为成分个数; y_i 为第 i 个成分, y_i 对应的取值称为分量,表示 i 成分所占的比例。

1.2 成分数据方法

1.2.1 ilr 方法

ilr 方法通过构建标准正交基将成分数据进行对数比转换,以使成分数据从单形空间 \mathcal{S}^D 映射到欧式空间 R^{D-1} ,再用经典统计方法分析。本研究选取最常用的标准正交基,即将感兴趣的成分放在第一位的标准正交基,以捕获该成分的所有相关信息。假设 $\mathbf{y}=(y_1, y_2, \dots, y_D) \in \mathcal{S}^D$, 为 D 元成分数据,等距对数比转换定义为(Egozcue et al., 2003; Fišerová et al., 2011):

$$\begin{aligned} Z_i &= \text{ilr}(y_i) \\ &= \sqrt{\frac{D-i}{D-i+1}} \ln \left(\frac{y_i}{\sqrt{\prod_{k=i+1}^D y_k}} \right) \\ &= \sqrt{\frac{1}{(D-i)(D-i+1)}} \left(\ln \frac{y_i}{y_{i+1}} + \ln \frac{y_i}{y_{i+2}} + \dots + \ln \frac{y_i}{y_D} \right) \end{aligned} \quad (2)$$

式(2)中, y_k 为第 k 个成分, $k = i + 1, i + 2, \dots, D; i = 1, 2, \dots, D-1$ 。ilr 将单形空间上的成分数据 $\mathbf{y}=(y_1, y_2, \dots, y_D)$ 变换为欧式空间上的普通向量 $\mathbf{Z}=(Z_1, Z_2, \dots, Z_{D-1})$, 并假设 \mathbf{Z} 服从多元正态分布,其联合密度函数为:

$$f_{\text{ilr}}(\mathbf{Z}; \mathbf{a}; \mathbf{b}) = \frac{1}{\left[\sqrt{(2\pi)^{D-1} |\mathbf{b}|} \right] \prod_{i=1}^D y_i} \exp \left[-\frac{1}{2} (\mathbf{Z} - \mathbf{a}) \mathbf{b}^{-1} (\mathbf{Z} - \mathbf{a})^t \right] \quad (3)$$

记为 $\mathbf{Z} \sim f_{\text{ilr}}(\mathbf{Z}; \mathbf{a}; \mathbf{b})$ 。式(3)中,向量 $\mathbf{a}=(a_1, a_2, \dots, a_k)$ 为 \mathbf{Z} 均值向量,满足 $a_k = E(Z_k)$; $\mathbf{b}=(b_{11}, b_{12}, \dots, b_{kl})$ 为 \mathbf{Z} 方差协方差阵,满足 $b_{kl} = \text{Cov}(Z_k, Z_l), k, l = 1, 2, \dots, D-1$ 。

似然函数(likelihood function)是指在参数给定条件下观察值的分布。ilr 似然函数为:

$$l_{\text{ilr}}(\mathbf{Z}; \mathbf{a}; \mathbf{b}) = \frac{1}{\left[\sqrt{(2\pi)^{D-1} |\mathbf{b}|} \right] \prod_{i=1}^D y_i} \exp \left[-\frac{1}{2} (\mathbf{Z} - \mathbf{a}) \mathbf{b}^{-1} (\mathbf{Z} - \mathbf{a})^t \right] \quad (4)$$

式(4)中符号或字母含义同式(3), c 表示已指定参

数 \mathbf{a} 和 \mathbf{b} 。

ilr 满足尺度缩放性和扰动不变性,根据 ilr 计算 \mathbf{y} 的均数 $E(\mathbf{y})$,可准确描述 \mathbf{y} 的集中趋势。均数 $E(y_i)$ 定义为:

$$E(y_i) = \text{ilr}^{-1} \left[E(Z_i) \right], i = 1, 2, \dots, D-1 \quad (5)$$

$$E(y_D) = 1 - \sum_{i=1}^{D-1} E(y_i) \quad (6)$$

因为单个成分的方差不包含成分间相互依赖的任何信息,而一个成分的变化必然受其他成分变化的影响,因此,成分数据的离散趋势用变异矩阵 \mathbf{T} 即成分间对数比方差进行描述,其中变异矩阵元素 t_{ij} 越接近 0,相应两个成分对数比方差越小,成分间比例(即共同依赖性)越强。

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1D} \\ t_{21} & t_{22} & \cdots & t_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ t_{D1} & t_{D2} & \cdots & t_{DD} \end{bmatrix}, t_{ij} = \text{Var} \left(\ln \frac{y_i}{y_j} \right), \quad i, j = 1, 2, \dots, D \quad (7)$$

1.2.2 Dirichlet 方法

Dirichlet 分布是一类在实数域以正单形为支撑集的高维连续概率分布,是 Beta 分布在高维情形的推广, Beta 分布是一个 2 维的 Dirichlet 分布 (Lin, 2016; Ng et al., 2011)。假设 D 元成分数据 $\mathbf{y} = (y_1, y_2, \dots, y_D) \in \mathcal{S}^D$ 服从 Dirichlet 分布,其密度函数 (Lin, 2016; Ng et al., 2011) 为:

$$f_{\text{Dir}}(\mathbf{y}; \boldsymbol{\alpha}) = \frac{1}{\beta(\boldsymbol{\alpha})} \prod_{i=1}^D y_i^{\alpha_i - 1} = \frac{\Gamma(\sum_{i=1}^D \alpha_i)}{\prod_{i=1}^D \Gamma(\alpha_i)} \prod_{i=1}^D y_i^{\alpha_i - 1} \quad (8)$$

记为 $\mathbf{y} \sim f_{\text{Dir}}(\mathbf{y}; \boldsymbol{\alpha})$ 。式(8)中, β 为多元的 Beta 函数; 向量 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_D)$ 为无量纲分布参数,满足 $\alpha_i > 0$; Γ 为伽马函数。

Dirichlet 似然函数定义为:

$$l_{\text{Dir}}(\mathbf{y}; \boldsymbol{\alpha}) = \frac{1}{\beta(\boldsymbol{\alpha})} \prod_{i=1}^D y_i^{\alpha_i - 1} = \frac{\Gamma(\sum_{i=1}^D \alpha_i)}{\prod_{i=1}^D \Gamma(\alpha_i)} \prod_{i=1}^D y_i^{\alpha_i - 1} \quad (9)$$

式(9)中符号或字母含义同式(8), c 表示已指定参数 $\boldsymbol{\alpha}$ 。

基于 Dirichlet 分布计算成分数据各成分的均数 $E(y_i)$ 和标准差 $\sigma(y_i)$,可定量描述成分数据各成分的集中趋势和离散趋势:

$$E(y_i) = \frac{\alpha_i}{\sum_{i=1}^D \alpha_i}, i = 1, 2, \dots, D \quad (10)$$

$$\sigma(y_i) = \sqrt{\frac{\alpha_i \left[\left(\sum_{i=1}^D \alpha_i \right) - \alpha_i \right]}{\left(\sum_{i=1}^D \alpha_i \right)^2 \left[\left(\sum_{i=1}^D \alpha_i \right) + 1 \right]}} = \sqrt{\frac{E(y_i) \left[1 - E(y_i) \right]}{\left(\sum_{i=1}^D \alpha_i \right) + 1}}, i = 1, 2, \dots, D \quad (11)$$

1.2.3 FD 方法

FD 分布是指从多个独立等比例伽马随机变量基开始,给第 i 个元素随机分配一个独立的伽马随机变量,构成 D 维 ($D > 1$) 正相关随机变量,最后将 D 维正相关随机

变量基归一化处理。假设 D 元成分数据 $\mathbf{y} = (y_1, y_2, \dots, y_D) \in \mathcal{S}^D$ 服从 FD 分布,FD 分布可看作多个 Dirichlet 分布的有限混合,通过对 Dirichlet 重新参数化,重新定义均向量结构为单形空间中各成分重心与第 i 个顶点 ($i = 1, 2, \dots, D$) 的线性凸组合,构建各成分均向量不同但精度参数相同的 Dirichlet 混合分布,即 FD 分布,其密度函数 (Migliorati et al., 2017a; Ongaro et al., 2013) 为:

$$f_{\text{FD}}(\mathbf{y}; \boldsymbol{\alpha}; \mathbf{p}; \tau) = \frac{\Gamma \left\{ \left[\left(\sum_{i=1}^D \alpha_i \right) + \tau \right] + \tau \right\}}{\prod_{i=1}^D \Gamma(\alpha_i)} \left(\prod_{i=1}^D y_i^{\alpha_i - 1} \right) \sum_{i=1}^D \left(p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i + \tau)} y_i^\tau \right) = \sum_{i=1}^D p_i f_{\text{Dir}}(\mathbf{y}, \alpha_i) \quad (12)$$

记为 $\mathbf{y} \sim f_{\text{FD}}(\mathbf{y}; \boldsymbol{\alpha}; \mathbf{p}; \tau)$ 。式(12)中,向量 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_D)$ 为形状参数,满足 $\alpha_i > 0$ 。当所有 $\alpha_i < 1$ 时,FD 分布密度函数为 U 形,当所有 $\alpha_i = 1$ 时,FD 分布密度函数为常数,当所有 $\alpha_i > 1$ 时,FD 分布密度函数为 $(\alpha_i - 1) / \left[\left(\sum_{i=1}^D \alpha_i \right) - D \right]$ 的单峰模式 (unimodality); 向量 $\mathbf{p} = (p_1, p_2, \dots, p_D)$ 为概率参数,满足 $0 \leq p_i < 1, \sum_{i=1}^D p_i = 1; \tau > 0, \tau$ 越大,第 i 成分的模式 (mode) 向单形第 i 个顶点移动越多; $\boldsymbol{\alpha} = \boldsymbol{\alpha} + \boldsymbol{\tau} e_i$, e_i 表示除第 i 位置取值为 1 外,其余位置取值为 0 的向量。 Γ 为伽马函数。

FD 似然函数为:

$$l_{\text{FD}}(\mathbf{y}; \boldsymbol{\alpha}; \mathbf{p}; \tau) = \sum_{i=1}^D p_i f_{\text{Dir}}(\mathbf{y}, \alpha_i) \quad (13)$$

式(13)中符号或字母含义同式(12), c 表示已指定参数 $\boldsymbol{\alpha}, \mathbf{p}, \tau$ 。但需注意 FD 为混合结构模型,在估计最大似然函数时作为一个数据缺失问题处理,FD 对数似然函数为:

$$\log [l_{\text{FD}}(\mathbf{y}; \boldsymbol{\alpha}; \mathbf{p}; \tau)] = \sum_{j=1}^n \sum_{i=1}^D m_{ji} \log [p_i f_{\text{Dir}}(\mathbf{y}, \alpha_i)] \quad (14)$$

式(14)中, $\mathbf{m}_j = (m_{j1}, m_{j2}, \dots, m_{jD})$ 表示缺失数据,取值为 0 或 1,当第 j 个观测在第 i 个成分混合模型出现时, $m_{ji} = 1$, 否则为 0。 n 为样本量; $j = 1, 2, \dots, n; i = 1, 2, \dots, D$ 。其余符号或字母含义同式(12)。FD 分布允许多峰模式 (multimodality),选择合适的 α_i, p_i 后,通过增加 τ 的值可以达到 k 峰模式 ($k \leq D$)。

基于 FD 分布计算成分数据各成分的均数 $E(y_i)$ 和标准差 $\sigma(y_i)$,可用于描述成分数据各成分的集中趋势和离散趋势:

$$E(y_i) = \frac{\alpha_i + p_i \tau}{\left(\sum_{h=1}^D \alpha_h \right) + \tau}, i = 1, 2, \dots, D \quad (15)$$

$$\sigma(y_i) =$$

$$\sqrt{\frac{E(y_i) \left[1 - E(y_i) \right]}{\left(\sum_{h=1}^D \alpha_h \right) + \tau + 1} + \frac{\tau^2 p_i (1 - p_i)}{\left[\left(\sum_{h=1}^D \alpha_h \right) + \tau \right] \left[\left(\sum_{h=1}^D \alpha_h \right) + \tau + 1 \right]}}, i = 1, 2, \dots, D \quad (16)$$

1.2.4 成分数据方法优缺点及差异

上述 3 种方法存在差异,各有优缺点(Di Brisco et al.,

2017; Migliorati et al., 2017a; Ng et al., 2011; Ongaro et al., 2013)(表 1)。

表 1 3 种成分数据方法优缺点及差异

Table 1 The Advantages, Disadvantages and Differences of Three Compositional Data Methods

方法	优点	缺点	差异
ilr	消除成分数据的定和限制; 保持变换前后向量的加法、数乘、内积运算结果不变; 计算简便	不满足子成分一致性和置换不变性; 依赖于标准正交基的选取且不唯一, 导致结果不唯一; 结果解释不直观	通过构建标准正交基将成分数据进行对数比转换; 采用变异矩阵描述成分数据的离散趋势
Dirichlet	满足置换不变性和子成分一致性; 参数易解释	不适用于某些组内弱正相关的成分数据; 对于成分间只含强相关性的成分数据适用性较弱	Dirichlet 分布是 Beta 分布在多变量上的推广, 可直接拟合和为 1 的多个频率
FD	满足置换不变性和子成分一致性; 简单且清晰的解释成分; 通过聚类分析将人群分为不同集群	相关结构完全为负	FD 分布是多个 Dirichlet 分布的有限混合

1.2.5 成分数据方法拟合效果评估

用 AIC 准则(Akaike information criterion, AIC)和 BIC 准则(Bayesian information criterion, BIC)评价 ilr、Dirichlet 和 FD 3 种方法拟合成分数据特征效果, AIC 和 BIC 越小, 描述成分数据越准确, 拟合效果更好。

$$AIC = -2 \log(\text{likelihood}) + 2k \quad (17)$$

$$BIC = -2 \log(\text{likelihood}) + \log(n)k \quad (18)$$

式(17)和(18)中, *likelihood* 为似然函数; *k* 为参数个数; *n* 为样本量。

1.3 成分数据方法的应用

在 PubMed 和中国知网检索 2015 年 1 月—2020 年 10 月的 24 h 活动行为时间相关研究, 英文检索关键词为“24 hour”“physical activity”“compositional data”, 中文检索关键词为“身体活动”“行为活动”“24 h 行为”“成分数据”。最终共筛选出符合条件的 24 h 行为时间相关研究文献 86 篇(英文 85 篇, 中文 1 篇), 包括 2018 年以前 32 篇, 2019 年 27 篇, 2020 年 26 篇。其中, 85 项研究采用 ilr 方法将 24 h 行为活动成分数据视为解释变量, 研究其对各类健康结局的影响。研究显示, 一天 24 h 限制条件下, MVPA、LPA、SB 或 SLP 与身体质量指数(Curtis et al., 2020)、体脂率(Pelclová et al., 2020)、脂肪指数(Oviedo-Caro et al., 2020)等肥胖指标、心脏代谢风险评分(Lee et al., 2020; Swindell et al., 2020)、心肺适应性(Oviedo-Caro et al., 2020)以及健康相关的生活质量关联(Curtis et al., 2020), 将 MVPA 时间替换为其他行为(LPA、SB 或 SLP), 将增加肥胖指标和心脏代谢风险、降低心肺适应性以及健康相关的生活质量。此外, 增加 MVPA 也可能使老年人的骨骼健康得到改善(Rodríguez-Gómez et al., 2020)。另一方面, SB 与抑郁症状存在正关联, 将 SB 时间分配给 MVPA 或 SLP 可缓解抑郁症状(Del Pozo Cruz et al., 2020)。SB

与全因死亡率也存在正关联, LPA 与全因死亡率负关联, 将 MVPA 或 LPA 时间替换成 SB 时间会增加全因死亡率(Von Rosen et al., 2020)。截至 2020 年 10 月, 仅有 1 项研究将 24 h 行为活动成分数据视作结局变量, 采用 Dirichlet 方法分析儿童视屏时间与 24 h 运动行为之间的关系, 结果表明, 与 2~3 岁时每天注视屏幕 1 h 或更少的儿童相比, 2~3 岁时每天注视屏幕 3 h 或更长时间的儿童在 5.5 岁时表现出更多的 SB, 较少的 MVPA 和 LPA(Chen et al., 2020)。作为最新提出的成分数据方法, FD 理论在 24 h 时间使用流行病学研究领域中仍未受到关注, 但方法学研究和实例结果均表明 FD 方法的灵活性使其比 ilr 和 Dirichlet 方法更能捕获橄榄油或者沉淀物成分的特征(Migliorati et al., 2017a; Ongaro et al., 2013)。因此, 采用 FD 模型将 24 h 活动行为成分数据视为结局变量并直接拟合, 以及基于 FD 模型聚类进行 24 h 活动行为模式识别, 仍是方法学应用的创新领域, 尚需更多的探索。

1.4 成分数据的可视化

与传统数据不同, 成分数据中各成分相互约束, 可用三元图进行可视化。其中, 三元图是各边长度均为 1 的等边三角形, 3 个顶点分别表示 3 个成分。离顶点越远, 顶点所在成分的分量越小, 反之越大。由点出发向各边零坐标正方向作平行线, 平行线与成分所在边的交点到零坐标的距离即为该点在该成分的分量。成分 A、B、C 所在三角形边分别为 ZX、XY、YZ, O 点在成分 A、B、C 的分量分别为 ZA=0.4, XB=0.4, YC=0.2(图 1)。

成分数据均向量作为一个点描绘在三元图上代表整个成分数据的集中趋势。三元图上的等密度水平线反映成分数据各成分离散程度, 且等密度水平线越密集, 对应成分离散程度越小, 反之越大。

当成分数据中包含 3 个以上成分时(如 MVPA-LPA-

SB-SLP),可分别绘制其中3个子成分的三元图。

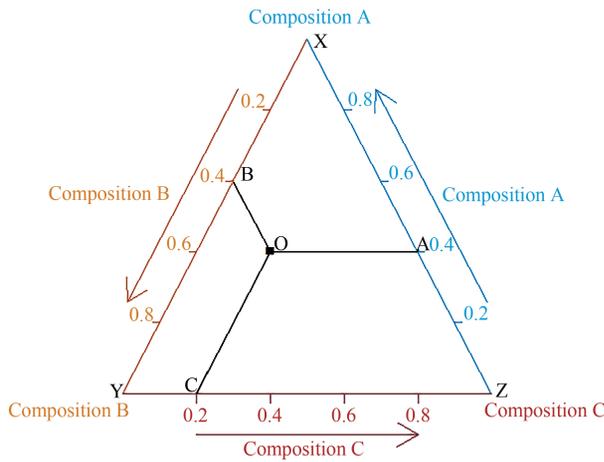


图1 三元图示例

Figure 1. An Example of Ternary Graph

1.5 统计软件

本研究使用 R 3.6.0 软件完成,其中,ilr 和 Dirichlet 方法相关 R 包为 compositions 包 (Van den Boogaart et al., 2018),FD 方法相关 R 包为 FlexDir 包 (Migliorati et al., 2017b)。

2 实例分析

2.1 研究对象与方法

本研究实例数据来自 2005—2006 年美国健康与营养

调查 (National Health and Nutrition Examination Survey, NHANES) 的 16~85 岁 ($n=3\ 682$) 美国居民日常 24 h 活动行为,身体活动采用加速度计 (ActiGraph AM 7164) 测量获得。研究人群中男性 1 823 人,女性 1 859 人;青年人 (16~44 岁) 1 859 人、中年人 (45~59 岁) 733 人、老年人 (60~85 岁) 1 090 人。本研究在 CoDA 框架下,将 ilr、Dirichlet 和 FD 方法应用于 NHANES 数据,分析 24 h 活动行为时间分布特征,用 AIC 和 BIC 准则评价 3 种方法拟合效果,比较 3 种方法在人群 24 h 活动行为时间成分数据的适用性。选择最优的拟合方法计算人群活动行为时间比例的均数和变异矩阵 (ilr) 或者标准差 (Dirichlet 和 FD),描述人群 24 h 活动行为时间比例的集中趋势和离散趋势,并采用对数似然比检验 (log likelihood ratio test, LLRT) 比较活动行为时间分布的性别和年龄差异。

2.2 结果

2.2.1 研究人群 24 h 活动行为时间使用分布特征

采用 ilr、Dirichlet 和 FD 方法分析人群 24 h 活动行为时间分布特征 (表 2~表 6)。

本研究发现 (表 2~表 3),研究人群平均花费在 MVPA、LPA、SB 和 SLP 的时间分别为 34.32、398.58、590.44、416.66 min。无论是分性别还是分年龄的拟合优度结果均表明,ilr 方法的 AIC 和 BIC 较小,更适合拟合研究人群 24 h 活动行为时间分布。总人群、男性、女性、青年人、中年人和老年人的 ilr 结果显示,一天中各人群用于 SB 时间的比例最大,其次为睡眠。

表 2 不同性别研究人群 24 h 活动行为时间使用分布特征
Table 2 Distribution Characteristics of 24 h Movement Time-Use in Different Gender Groups

	MVPA ($M \pm SD$)	LPA ($M \pm SD$)	SB ($M \pm SD$)	SLP ($M \pm SD$)	AIC	BIC
总人群						
绝对量/min	34.32±32.78	398.58±122.37	590.44±136.40	416.66±84.83		
相对量/%						
ilr ^a	1.44	27.44	41.61	29.51	-43 132.39	-43 076.49
Dirichlet	2.63±2.71	27.35±7.56	40.72±8.33	29.30±7.72	-37 379.57	-37 354.73
FD	2.36±2.26	27.66±8.63	40.92±9.14	29.06±6.74	-37 983.78	-37 934.09
男性						
绝对量/min	42.88±37.82	396.27±128.45	586.84±147.66	414.01±84.27		
相对量/%						
ilr ^a	1.89	27.27	41.35	29.49	-20 393.87	-20 344.29
Dirichlet	3.22±3.17	27.17±7.99	40.36±8.81	29.25±8.17	-17 368.19	-17 346.16
FD	2.88±2.59	27.55±9.25	40.64±9.77	28.93±7.01	-17 738.79	-17 694.72
女性						
绝对量/min	25.93±24.16	400.84±116.08	593.98±124.31	419.26±85.32		
相对量/%						
ilr ^a	1.11	27.56	41.83	29.50	-23 011.01	-22 961.26
Dirichlet	2.12±2.26	27.49±6.99	41.06±7.70	29.33±7.13	-20 248.68	-20 226.57
FD	1.91±1.89	27.75±7.92	41.21±8.41	29.13±6.28	-20 525.43	-20 481.20

注:a表示 ilr 方法采用变异矩阵描述成分数据离散趋势,因此这里仅列出 ilr 计算成分数据的均值;下同。

表 3 不同年龄研究人群 24 h 活动行为时间使用分布特征

Table 3 Distribution Characteristics of 24 h Movement Time-Use in Different Age Groups

	MVPA ($M \pm SD$)	LPA ($M \pm SD$)	SB ($M \pm SD$)	SLP ($M \pm SD$)	AIC	BIC
青年人						
绝对量/min	44.60±34.65	415.77±117.62	557.96±124.81	421.68±81.53		
相对量/%						
ilr ^a	2.36	28.76	39.11	29.77	-21 530.79	-21 481.04
Dirichlet	3.39±2.83	28.60±7.07	38.45±7.62	29.56±7.14	-18 592.28	-18 570.16
FD	3.12±2.35	28.87±8.20	38.69±8.54	29.32±6.14	-18 999.51	-18 955.28
中年人						
绝对量/min	34.24±29.21	430.90±116.22	581.05±127.86	393.81±81.49		
相对量/%						
ilr ^a	1.66	29.86	40.81	27.67	-8 781.74	-8 740.37
Dirichlet	2.69±2.53	29.66±7.13	40.08±7.65	27.57±6.98	-7 635.29	-7 616.90
FD	2.44±2.10	29.93±8.18	40.28±8.52	27.35±6.07	-7 757.25	-7 720.48
老年人						
绝对量/min	16.85±22.89	347.52±118.80	652.16±140.18	423.47±89.87		
相对量/%						
ilr ^a	0.56	23.50	46.08	29.86	-14 399.26	-14 354.31
Dirichlet	1.63±2.13	23.67±7.17	45.02±8.39	29.68±7.70	-12 486.36	-12 466.38
FD	1.39±1.71	24.12±8.28	45.08±9.07	29.41±6.65	-12 724.77	-12 684.82

表 4 不同性别、年龄研究人群 24 h 活动行为时间使用分布差异检验

Table 4 Test of the Differences in the Distribution of 24 h Movement Time-Use among Different Gender and Age Groups

比较组别	统计量	P
男性 vs 女性	2 617.14	<0.001
青年人 vs 中年人	12 749.05	<0.001
青年人 vs 老年人	7 131.53	<0.001
中年人 vs 老年人	5 617.52	<0.001

表 5 不同性别研究人群 24 h 活动行为时间使用比例变异矩阵

Table 5 Variation Matrix of Ratio for 24 h Movement Time-Use in Different Gender Groups

	MVPA	LPA	SB	SLP
总人群				
MVPA	0.00	1.39	1.97	1.70
LPA	1.39	0.00	0.30	0.21
SB	1.97	0.30	0.00	0.15
SLP	1.70	0.21	0.15	0.00
男性				
MVPA	0.00	1.23	1.92	1.58
LPA	1.23	0.00	0.35	0.22
SB	1.92	0.35	0.00	0.17
SLP	1.58	0.22	0.17	0.00
女性				
MVPA	0.00	1.39	1.87	1.67
LPA	1.39	0.00	0.25	0.20
SB	1.87	0.25	0.00	0.13
SLP	1.67	0.20	0.13	0.00

表 6 不同年龄研究人群 24 h 活动行为时间使用比例变异矩阵

Table 6 Variation Matrix of Ratio for 24 h Movement Time-Use in Different Age Groups

	MVPA	LPA	SB	SLP
青年人				
MVPA	0.00	0.72	0.92	0.80
LPA	0.72	0.00	0.25	0.17
SB	0.92	0.25	0.00	0.13
SLP	0.80	0.17	0.13	0.00
中年人				
MVPA	0.00	0.90	1.18	1.02
LPA	0.90	0.00	0.23	0.18
SB	1.18	0.23	0.00	0.15
SLP	1.02	0.18	0.15	0.00
老年人				
MVPA	0.00	1.87	2.63	2.33
LPA	1.87	0.00	0.33	0.24
SB	2.63	0.33	0.00	0.15
SLP	2.33	0.24	0.15	0.00

对数似然比检验结果显示,男性、女性 24 h 活动行为时间比例分布存在差异(表 4),男性 MVPA 时间比女性高,SB 时间比女性低(表 2)。不同年龄段人群的 24 h 活动行为时间比例分布不同,相比于青年人,中年人和老年人 MVPA 时间比例更低,SB 时间比例更高(表 3)。

不同性别和不同年龄研究人群 24 h 活动行为变异矩阵结果显示,各人群 MVPA 时间比例均与其他行为时

间依赖性较小,SB、SLP和LPA时间三者的依赖性较强(表5~表6)。

2.2.2 研究人群24 h活动行为时间使用分布特征可视化

采用三元图对ilr方法计算研究人群24 h活动行为时

间比例分布结果进行可视化(图2)。

研究显示,总人群MVPA时间比例方向的等密度曲线相对其他行为方向更疏散,即MVPA时间比例离散程度更大。

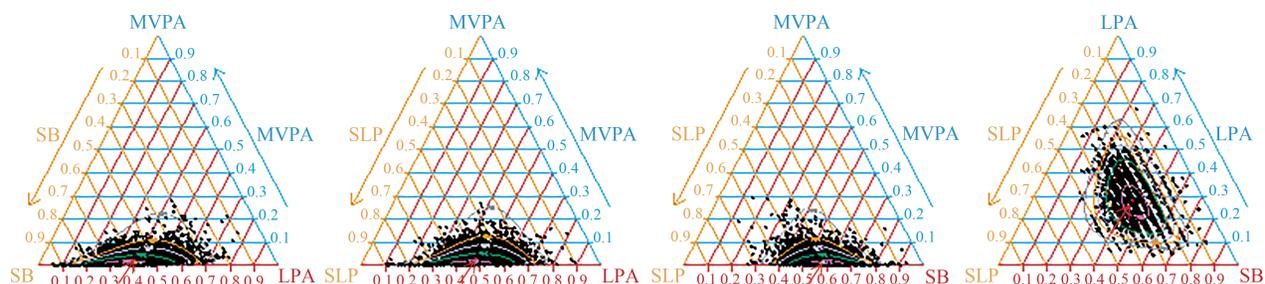


图2 ilr方法计算的总人群24 h活动行为时间使用比例三元图

Figure 2. Ternary Diagram of Ratio for 24 h Movement Time-Use in the Total Population by ilr Method

注:图中红色箭头指向的红点反映总人群24 h活动行为时间比例整体平均水平。

3 讨论

随着PA、SB和SLP健康效应研究的不断深入,证明除MVPA外,LPA和充足的睡眠也可以对人体健康产生益处(Chaput et al., 2016; Poitras et al., 2016),SB对多种健康指标都有不利的影响(Carson et al., 2016),且各行为间可以彼此缓和和健康的影响(Chaput et al., 2017)。仅考虑24 h活动行为连续谱中某个活动行为与健康的关系,限制了对活动行为间相互作用对健康影响的理解,24 h活动行为时间使用流行病学研究正成为当前公共卫生和运动科学等健康相关领域重点研究方向,但与此同时,将24 h行为时间使用数据视作成分数据,也给统计分析带来一定的挑战(Chaput et al., 2014; Dumuid et al., 2020),积极探索成分数据方法理论及应用对促进时间使用流行病学研究具有重要意义。

本研究介绍了ilr、Dirichlet和FD 3种成分数据方法及其优缺点。理论上,ilr方法消除了成分数据的定和限制,将成分数据从单形空间转换到欧式空间,避免成分数据建模时造成多重共线性问题,使得多元统计方法得以应用。ilr方法转换计算简便,成分均值以及变异矩阵可用于描述成分数据的集中趋势和离散趋势,但ilr方法在处理各种单形独立形式(如中立性)、重要的成分操作(如合并某些同质成分)存在一定困难。更重要的是,建模时ilr方法受标准正交基选取影响,分析结果不唯一,且模型参数解释原始成分变量不直观。此外,ilr方法采用变异矩阵描述成分内共同依赖性,变异矩阵元素取值越小,表明相应成分间共同依赖性越强,但无法明确相关方向。相比于ilr方法,Dirichlet和FD方法直接基于单形空间建模,结果更直观,采用建模后成分平均值±标准差描述成分数据的集中趋势和离散趋势。Dirichlet方法满足ilr方法不能满足的CoDA条件,即子成分一致性以及置换不变性等。由于Dirichlet成分隐含强独立结构、参数易解释,在

很多情况下,Dirichlet被作为成分间具有强独立关系的成分数据建模标准参考,但Dirichlet的成分强独立结构、相关结构完全为负性质,不适于组分间弱正相关或强相关的成分数据建模。相比于Dirichlet方法,FD方法为Dirichlet方法的推广,在建模依赖性方面具有更大的灵活性,可通过配置不同参数适应各种独立形式的成分数据,满足成分数据更多的分析要求,更能捕获成分数据特征。但FD方法实际上是一种混合结构模型,具有强混合成分(集群)链接特点,通常的单模态参数化模型过于严格,导致未知的推理复杂度。FD方法因相关结构完全为负,不允许在成分之间建立正相关模型。因此,从理论上讲,3种方法在分析成分数据时各有优缺点,需结合成分数据实际情况选择合适的分析方法。实际应用中,成分数据方法拟合效果会受到样本量、成分比例、成分方差、成分相关性等因素的影响,模拟研究显示,在成分数据具有大样本、成分间比例相差越大、成分方差越小、正相关性越强情况下,ilr方法有更好的拟合效果(刘晴,2020)。而对于Dirichlet和FD模型适应的成分数据特征,需要后续更多的模拟研究进行探讨。

本研究将3种成分数据方法应用于NHANES人群的24 h活动行为时间的实例研究中,结果显示,该实例数据中ilr方法具有最好的拟合优度,即ilr方法能更准确描述人群的24 h活动行为时间分布特征。但鉴于目前鲜有研究采用Dirichlet或FD方法分析24 h活动行为特征和各类健康效应,3种成分数据方法适用条件仍在理论探索阶段,后续可根据ilr、Dirichlet和FD原理进行成分数据模拟研究,进一步探讨3种成分数据方法适用条件。ilr方法结果显示,16~85岁美国人群一天中MVPA时间为20.74 min,低于直接计算的绝对量34.32 min,也低于《美国身体活动指南》成年人平均每日MVPA最低推荐量32.14 min(Piercy et al., 2018)。16~85岁美国人群一天SB为599.18 min,略

高于直接计算的绝对量 590.44 min, 但远高于 2007—2016 年基于问卷数据计算的 NHANES 人群的加权平均 SB (342.00~474.00 min) (Du et al., 2019), 目前尚未有成人 SB 具体推荐量, 但《美国身体活动指南》建议多动少坐。现有研究表明, 以《美国身体活动指南》的成人 PA 推荐量为参考, 美国成人有氧和加强肌肉的 PA 达标率呈上升趋势, 但变化不明显, 总体达标率仍然较低 (Du et al., 2019; Whitfield et al., 2019)。美国成人的 SB 时间较以前明显增加 (Du et al., 2019)。另外, 基于传统统计方法的研究表明, 男性比女性更倾向于运动, 在男性中, 中年人更容易久坐; 女性中, 老年人更容易久坐 (Azevedo et al., 2007)。但基于成分数据方法研究发现, 男性比女性更不活跃, 工作中年龄更小的成年人久坐时间更长 (Gupta et al., 2018)。本研究发现, 不同性别、年龄的人群 24 h 活动行为时间使用分布存在一定差别, 且男性的 MVPA 时间比女性长, SB 时间比女性短。相比于青、中年人, 老年人的 MVPA 时间更短, SB 时间更长。研究结果不一致的原因可能与统计分析方法、数据收集方式及时间等因素有关, 但研究均表明年龄和性别是 PA、SB 的两个潜在重要决定因素。本研究采用的是 2005—2006 年 NHANES 数据, 时效性较差, 但该年的身体活动数据是基于加速度计收集的。在人体活动测量中, 加速度计较于问卷更具客观、精确、实用 (向剑锋等, 2015; 赵壮壮等, 2015; Dyrstad et al., 2014), 将其获得数据用于方法学实例研究结果更可靠。

4 结论

ilr、Dirichlet 和 FD 方法理论上各有优缺点, 需结合成分数据实际情况选择合适的成分数据方法分析。本研究将 3 种方法应用于 2005—2006 年美国 NHANES 人群中分析 24 h 活动行为时间比例数据, 发现 ilr 有最好的拟合优度。未来的研究仍需更多模拟和应用研究, 探索各种成分数据方法适应的数据特征, 以促进时间使用流行病学研究, 对制定人体 24 h 活动指南、提升人体体质健康水平有着重要意义。

参考文献:

- 郭强, 汪晓赞, 蒋健保, 2017. 我国儿童青少年身体活动与久坐行为模式特征的研究[J]. 体育科学, 37(7): 17-29.
- 刘晴, 2020. 基于成分数据方法探讨儿童青少年体成分特征及影响因素[D]. 广州: 广东药科大学: 72-79.
- 宋俊辰, 李红娟, 王政淞, 2020. 时间使用流行病学在身体活动研究领域的应用[J]. 体育科学, 40(1): 79-88.
- 卫文, 2018. 全球 14 亿成年人身体活动不足[J]. 家庭医学, (10): 32.
- 向剑锋, 李之俊, 2015. 加速度计和体力活动日记监测日常体力活动的效度研究[J]. 中国体育科技, 51(6): 128-133.
- 闫静弋, 张俊黎, 姜迎, 等, 2019. 2010—2012 年山东省成年居民睡眠状况[J]. 卫生研究, 48(6): 884-887.
- 于洪军, 仇军, 2013. 身体活动负荷对我国老年人患慢性疾病风险率

的影响研究: 基于对清华大学老年人人群 PASE 问卷的流行病学调查[J]. 中国体育科技, 49(2): 139-145.

- 张尧庭, 2000. 成分数据统计分析引论[M]. 北京: 科学出版社: 1-83.
- 赵壮壮, 徐京朝, 2015. Actigraph 及其信效度研究[J]. 体育成人教育学报, 31(1): 30-35.
- 周志雄, 2015. 30~45 岁男性久坐少动行为与心血管疾病危险因素的关系及干预[D]. 北京: 北京体育大学: 26-38.
- AZEVEDO M R, ARAÚJO C L, REICHERT F F, et al., 2007. Gender differences in leisure-time physical activity [J]. *Int J Public Health*, 52(1): 8-15.
- BAUMAN A, AINSWORTH B E, SALLIS J F, et al., 2011. The descriptive epidemiology of sitting: A 20-country comparison using the international physical activity questionnaire (IPAQ) [J]. *Am J Prev Med*, 41(2): 228-235.
- CARSON V, HUNTER S, KUZIK N, et al., 2016. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: An update [J]. *Appl Physiol Nutr Metab*, 41[6 (Suppl 3)]: S240-S265.
- CHAPUT J P, CARSON V, GRAY C E, et al., 2014. Importance of all movement behaviors in a 24 hour period for overall health [J]. *Int J Environ Res Public Health*, 11(12): 12575-12581.
- CHAPUT J P, GRAY C E, POITRAS V J, et al., 2016. Systematic review of the relationships between sleep duration and health indicators in school-aged children and youth [J]. *Appl Physiol Nutr Metab*, 41[6 (Suppl 3)]: S266-S282.
- CHAPUT J P, SAUNDERS T J, CARSON V, 2017. Interactions between sleep, movement and other non-movement behaviours in the pathogenesis of childhood obesity [J]. *Obes Rev*, 18(S1): 7-14.
- CURTIS R G, DUMUID D, OLDS T, et al., 2020. The association between time-use behaviors and physical and mental well-being in adults: A compositional isotemporal substitution analysis [J]. *J Phys Act Health*, 17(2): 197-203.
- DEL POZO CRUZ B, ALFONSO-ROSA R M, MCGREGOR D, et al., 2020. Sedentary behaviour is associated with depression symptoms: Compositional data analysis from a representative sample of 3233 US adults and older adults assessed with accelerometers [J]. *J Affect Disord*, 265: 59-62.
- DI BRISCO A M, MIGLIORATI S, 2017. A special Dirichlet mixture model for multivariate bounded responses [C]. Milan, Italy: International Conference of the Classification and Data Analysis Group: 285-290.
- DU Y, LIU B, SUN Y, et al., 2019. Trends in adherence to the physical activity guidelines for Americans for aerobic activity and time spent on sedentary behavior among US adults, 2007 to 2016 [J]. *JAMA Netw Open*, doi: 10.1001/jamanetworkopen.2019.7597.
- DUMUID D, PEDIŠIĆ Ž, PALAREA-ALBALADEJO J, et al., 2020. Compositional data analysis in time-use epidemiology: What, why, how [J]. *Int J Environ Res Public Health*, doi: 10.3390/ijerph17072220.
- DYRSTAD S M, HANSEN B H, HOLME I M, et al., 2014. Comparison of self-reported versus accelerometer-measured physical activity [J]. *Med Sci Sports Exerc*, 46(1): 99-106.
- EGOZCUE J J, PAWLOWSKY-GLAHN V, MATEU-FIGUERAS G, et al., 2003. Isometric logratio transformations for compositional data analysis [J]. *Math Geol*, 35(3): 279-300.
- FIŠEROVÁ E, HRON K, 2011. On the Interpretation of orthonormal

- coordinates for compositional data[J]. *Math Geosci*, 43(4): 455-468.
- GUPTA N, MATHIASSEN S E, MATEU-FIGUERAS G, et al., 2018. A comparison of standard and compositional data analysis in studies addressing group differences in sedentary behavior and physical activity[J]. *Int J Behav Nutr Phys Act*, doi: 10.1186/s12966-018-0685-1.
- KEYES K M, MASLOWSKY J, HAMILTON A, et al., 2015. The great sleep recession: Changes in sleep duration among US adolescents, 1991—2012[J]. *Pediatrics*, 135(3): 460-468.
- LEE J, WALKER M E, MATTHEWS K A, et al., 2020. Associations of physical activity and sleep with cardiometabolic risk in older women[J]. *Prev Med Rep*, doi: 10.1016/j.pmedr.2020.101071.
- LIN J, 2016. On the Dirichlet distribution[D]. Kingston, Canada: Queen's University: 3-31.
- MIGLIORATI S, ONGARO A, MONTI G S, 2017a. A structured Dirichlet mixture model for compositional data: Inferential and applicative issues[J]. *Stat Comput*, 27(4): 963-983.
- MIGLIORATI S, BRISCO A M D, VESTRUCCI M, 2017b. FlexDir: Tools to Work with the Flexible Dirichlet Distribution[EB/OL]. <https://cran.r-project.org/web/packages/FlexDir/index.html>.
- NG W N, TIAN G L, TANG M L, 2011. Dirichlet and Related Distributions: Theory, Methods and Applications[M]. UK: John Wiley & Sons Ltd. : 37-87.
- ONGARO A, MIGLIORATI S, 2013. A generalization of the Dirichlet distribution[J]. *J Multivariate Anal*, 114: 412-426.
- OVIEDO-CARO M Á, BUENO-ANTEQUERA J, MUNGUÍA-IZQUIERDO D, 2020. Associations of 24-hours activity composition with adiposity and cardiorespiratory fitness: The PregnActive project[J]. *Scand J Med Sci Sports*, 30(2): 295-302.
- PATTERSON R, MCNAMARA E, TAINIO M, et al., 2018. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: A systematic review and dose response meta-analysis[J]. *Eur J Epidemiol*, 33(9): 811-829.
- PEDIŠIĆ Ž, 2014. Measurement issues and poor adjustments for physical activity and sleep undermine sedentary behaviour research: The focus should shift to the balance between sleep, sedentary behaviour, standing and activity[J]. *Kinesiology*, 46(1): 135-146.
- PEDIŠIĆ Ž, DUMUID D, OLDS T S, 2017. Integrating sleep, sedentary behaviour, and physical activity research in the emerging field of time-use epidemiology: Definitions, concepts, statistical methods, theoretical framework, and future directions[J]. *Kinesiology*, 49(2): 252-269.
- PELCLOVÁ J, ŠTEFELOVÁ N, DUMUID D, et al., 2020. Are longitudinal reallocations of time between movement behaviours associated with adiposity among elderly women? A compositional isotemporal substitution analysis[J]. *Int J Obes*, 44(4): 857-864.
- PIERCY K L, TROIANO R P, BALLARD R M, et al., 2018. The Physical Activity Guidelines for Americans[J]. *JAMA*, 320(19): 2020-2028.
- POITRAS V J, GRAY C E, BORGHESE M M, et al., 2016. Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth[J]. *Appl Physiol Nutr Metab*, 41[6 (Suppl 3)]: S197-S239.
- RODRÍGUEZ-GÓMEZ I, MAÑAS A, LOSA-REYNA J, et al., 2020. Prospective changes in the distribution of movement behaviors are associated with bone health in the elderly according to variations in their frailty levels[J]. *J Bone Miner Res*, 35(7): 1236-1245.
- SHEEHAN C M, FROCHEN S E, WALSEMANN K M, et al., 2019. Are US adults reporting less sleep? Findings from sleep duration trends in the National Health Interview Survey, 2004—2017[J]. *Sleep*, doi: 10.1093/sleep/zsy221.
- SWINDELL N, REES P, FOGELHOLM M, et al., 2020. Compositional analysis of the associations between 24-h movement behaviours and cardio-metabolic risk factors in overweight and obese adults with pre-diabetes from the PREVIEW study: Cross-sectional baseline analysis [J]. *Int J Behav Nutr Phys Act*, doi: 10.1186/s12966-020-00936-5.
- VAN DEN BOOGAART K G, TOLOSANA-DELGADO R, 2013. Analyzing Compositional Data with R[M]. New York: Springer-Verlag Berlin Heidelberg: 73-93.
- VAN DEN BOOGAART K G, TOLOSANA-DELGADO R, BRENN M, 2018. [EB/OL]. <https://rdrr.io/cran/compositions/man/compositions.html>.
- VON ROSEN P, DOHRN I M, HAGSTRÖMER M, 2020. Association between physical activity and all-cause mortality: A 15-year follow-up using a compositional data analysis[J]. *Scand J Med Sci Sports*, 30(1): 100-107.
- WHITFIELD G P, CARLSON S A, USSERY E N, et al., 2019. Trends in meeting physical activity guidelines among urban and rural dwelling adults: United States, 2008—2017[J]. *Morb Mortal Wkly Rep*, 68(23): 513-518.
- WHO, 2018. Physical Activity[EB/OL]. [2021-07-12]. <https://www.who.int/health-topics/physical-activity>.

(收稿日期:2020-09-07; 修订日期:2022-05-24; 编辑:尹航)