

·基础理论与方法·

降秩回归方法简介及其 Stata 程序开发

郑棒^{1,2,3} 刘琪^{1,2} 吕筠^{1,2} 余灿清^{1,2}

¹北京大学公共卫生学院流行病与卫生统计学系,北京 100191;²北京大学公众健康与重大疫情防控战略研究中心,北京 100191;³帝国理工学院公共卫生学院,英国伦敦 W6 8RP
通信作者:余灿清,Email: yucanqing@pku.edu.cn

【摘要】 降秩回归是一种用于降维的多变量线性回归,近年来被越来越多的应用于营养流行病学领域来提取膳食模式。但 Stata 软件作为主流的统计分析工具之一,目前尚无进行降秩回归的程序。本文介绍了降秩回归方法的基本思想、实现降秩回归的基本原理,并开发和公开发布了功能全面的 Stata 程序(rrr),最后结合中国慢性病前瞻性研究数据进行示例分析。rrr 程序的研发和公开,为该统计方法在流行病学和公共卫生领域的广泛应用提供了实用工具。

【关键词】 降秩回归; 流行病学; 应用; 软件实现

基金项目: 国家自然科学基金(81973125)

Introduction of reduced rank regression and development of a user-written Stata packageZheng Bang^{1,2}, Liu Qi¹, Lyu Jun¹, Yu Canqing¹

¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; ²Peking University Center for Public Health and Epidemic Preparedness & Response, Beijing 100191, China; ³School of Public Health, Imperial College London, London W6 8RP, UK
Corresponding author: Yu Canqing, Email: yucanqing@pku.edu.cn

【Abstract】 Reduced rank regression is an extended multivariate linear regression model with the function of dimension reduction. It has been more and more widely used in nutritional epidemiology research to understand people's dietary patterns in recent years. However, there has been no existing Stata package or command to implement reduced rank regression independently. Therefore, we developed a new user-written package named "rrr" for its implementation in Stata. This paper summarizes the methodology of reduced rank regression, the development and functions of the Stata rrr package and its application in the China Kadoorie Biobank dataset, with the aim of facilitating the future wide use of this statistical method in epidemiology and public health research.

【Key words】 Reduced rank regression; Epidemiology; Application; Software implementation

Fund program: National Natural Science Foundation of China (81973125)

降秩回归(reduced rank regression, RRR)是一种多变量统计方法,在使用多个自变量预测或解释多个因变量的同时达到降维的目的^[1-2]。降秩回归在国内外文献中大多特指降秩线性回归,即在传统多变量线性回归的基础上进行的拓展。传统多变量线性回归通过估计自变量的线性组合拟合多个

因变量,其线性组合的个数与因变量的个数相等。而降秩回归通过估计少于因变量个数的相互独立的自变量线性组合(降秩回归因子),最大程度解释因变量组的变异。

降秩回归近年来被广泛应用到了医学领域,尤其是营养流行病学中基于特定的因素提取膳食模

DOI:10.3760/cma.j.cn112338-20210222-00136

收稿日期 2021-02-22 本文编辑 李银鸽

引用格式:郑棒,刘琪,吕筠,等.降秩回归方法简介及其 Stata 程序开发[J].中华流行病学杂志,2022,43(3):403-408.

DOI: 10.3760/cma.j.cn112338-20210222-00136.

Zheng B, Liu Q, Lyu J, et al. Introduction of reduced rank regression and development of a user-written Stata package[J].

Chin J Epidemiol, 2022, 43(3):403-408. DOI: 10.3760/cma.j.cn112338-20210222-00136.



式(即各类食物摄入量的线性组合)^[3-5]。Hoffmann 等^[3]在欧洲肿瘤与营养前瞻性队列(European Prospective Investigation into Cancer and Nutrition)的 Potsdam 子队列中应用了该方法。他们使用 49 个食物组的摄入量数据预测 4 个糖尿病相关的营养素摄入指标(多不饱和脂肪酸与饱和脂肪酸摄入量比值、纤维素摄入量、镁摄入量、酒精摄入量),从而得到了一个与糖尿病患病风险有关联的膳食模式。目前国内使用降秩回归进行流行病学分析的研究较少。殷召雪等^[6]和程茅伟等^[7]分别用降秩回归得到了血压值(SBP 和 DBP)和代谢综合征生物标记物(腰围、SBP、DBP、TG、HDL-C、FPG)相关的膳食模式,并利用这些膳食模式研究食物摄入与疾病风险的关联。值得注意的是,本文所提及的膳食模式是指为了预测一组因变量(如生物学指标)而对一组膳食摄入自变量的综合(即线性组合),而不是通过对膳食摄入变量做因子分析而得到的隐变量。

一、降秩回归方法介绍

降秩回归是一种用于降维的多变量线性回归,适用于存在多个相关性较强的连续型因变量的情况。其在传统多变量线性回归的假设基础上,对回归系数矩阵或预测值矩阵的秩加以限制。假设自变量个数为 p (即自变量组: X_1, \dots, X_p), 因变量个数为 q (即因变量组: Y_1, \dots, Y_q), 样本量为 n , 自变量矩阵 $X(n \times p)$ 与因变量矩阵 $Y(n \times q)$ 均满秩, 且 $n > p \geq q$ 。传统的普通最小二乘估计(ordinary least squares, OLS)的目标是使残差平方和最小, 即求解回归系数矩阵 $B(p \times q)$, 使得 $L = \|Y - XB\|^2$ 最小。其中预测值矩阵用 \hat{Y}_{OLS} 表示, 即 $\hat{Y}_{OLS} = X\hat{B}_{OLS}$ 。

降秩回归方法对回归系数矩阵 $B(p \times q)$ 的秩进行限制, 要求在 B 的秩 $m \leq q$ 的前提下, 使残差平方和最小。这等同于求解 m 个不存在共线性的自变量线性组合(即 m 个降秩回归因子), 来解释因变量组的变异。该求解过程^[1-2, 8-9]可以通过对普通最小二乘法得到的预测值矩阵 \hat{Y}_{OLS} 做奇异值分解(或基于其协方差矩阵做主成分分析)来实现, 即 $\hat{Y}_{OLS} = U \sum V^T$ 。 m 个最优线性组合的估计值由提取出的前 m 个特征向量(V_m)获得, 即 $\hat{Y}_{OLS} V_m$, 而降秩回归的预测值矩阵 $\hat{Y}_{RRR} = \hat{Y}_{OLS} V_m V_m^T$ 。同样的, 降秩回归的回归系数矩阵的估计为 $\hat{B}_{RRR} = \hat{B}_{OLS} V_m V_m^T$ 。从主成分分析的角度来看, 降秩回归所得的 m 个自变量线性组合就是 OLS 估计出的预测值组的前 m 个主成分。

特别的, 在 $m=q$ 的情况下, 降秩回归能提

取出 q 个自变量线性组合但不能起到降维的作用, 此时其优点是这 q 个变量相互线性独立。而在规定 $m < q$ 的情况下, 对秩次的这一限制则能体现降秩回归的降维作用。在现实情况下, 使因变量组的预测残差平方和最小的秩次数为 q , 因为每减少一个秩次, 预测变量的数量(即可以使用的预测信息)就会减少, 对因变量组的拟合效果就会变差。因此降秩回归中对秩次 m 的选择是在降维和拟合度之间的取舍与平衡。此外, 降秩回归是多变量统计方法, 如果只有一个因变量($q=1$), 降秩回归的结果就等同于一般线性回归的结果。

二、降秩回归的软件实现与 Stata 程序开发

1. 降秩回归的软件实现: 目前, 主流的统计分析软件开发了降秩回归的程序或者宏。SAS 软件的偏最小二乘(PLS)过程或高性能偏最小二乘(HPLS)过程可以实现降秩回归(proc pls data=data method=RRR 或 proc hppls data=data method=RRR)^[10]。R 语言的 rrr 包^[11]和 rrpac 包^[12]也可以实现降秩回归。然而, Stata 软件目前尚无进行降秩回归的独立程序或命令, 仅有的 Stata 程序(PLSSAS)是通过调用 SAS 软件实现的^[13]。

基于以上情况, 笔者开发了用于实现降秩回归的 Stata 程序(rrr), 并提交到了波士顿大学“统计软件组件存档”[Statistical Software Components (SSC) archive]^[14]。读者可以通过在 Stata 软件里直接输入“ssc install rrr”语句安装, 该程序包含程序代码(rrr.ado)和帮助文件(rrr.sthlp)^[14]。

2. Stata 程序开发原理: Stata 程序 rrr 是按照降秩回归方法的基本思想进行原生开发, 即对 OLS 预测值矩阵 \hat{Y}_{OLS} 进行主成分分析。此处选择使用主成分分析(而不是奇异值分解)的算法去实现降秩回归是因为主成分分析可以基于 Stata 软件的官方 pca 命令来直接实现, 有助于提高程序的简洁性。另一方面是在 Stata 软件中做奇异值分解涉及到对矩阵的存储及运算, 会受到预测值矩阵 \hat{Y}_{OLS} 大小的限制(大样本时难以完成)^[15]。

其原理和步骤: ①利用 Stata 软件的 regress 命令, 分别用自变量组对每个因变量进行一般线性回归, 进而用 predict 语句得到线性预测值 \hat{Y}_{OLS} ; ②利用 pca 命令对这些预测值向量进行基于协方差矩阵的主成分分析, 使用 predict 语句得到前几个或所有主成分的估计值, 即降秩回归因子; ③使用 regress 命令计算提取出的降秩回归因子对原始因

变量的单独或累计方差解释百分比(R^2),以评价降秩回归的效果;并计算降秩回归因子对原始自变量的单独或累计方差解释百分比,以衡量其与自变量的相关性。

此外,分析中还可以使用 regress 命令,以提取的降秩回归因子对自变量组做线性回归,从而根据该模型的回归系数(即因子载荷)得到降秩回归因子的计算公式。

3. Stata 程序功能介绍: rrr 程序的主体语句形式是: rrr depvars [if] [in], x (indepvars) [rank() nostd savevar loadings]。其中因变量名列表和自变量名列表是必要语句,还可以选择是否需要筛选纳入样本、设定秩次、进行变量标准化、保存中间变量以及因子载荷输出(语句中的[]内的部分),详细说明见表 1。

程序运行后,Stata 软件在界面依次输出分析结果,包括 OLS 回归的参数估计结果、因变量 OLS 预测值的描述性统计、主成分分析结果、降秩回归因子对原始变量的方差解释百分比以及因子载荷等信息。

与此同时,程序将分析过程中新生成的变量保存在原始数据集中,包括提取的降秩回归因子、标准化后的自变量和因变量、因变量的 OLS 预测值等变量,具体变量解释见表 2。

三、降秩回归 Stata 程序的实例应用

本文利用中国慢性病前瞻性研究(China Kadoorie Biobank)^[16]第二次重复调查的膳食数据作为实例,介绍降秩回归和 rrr 程序的应用。此研究纳入 25 041 名调查对象(平均年龄为 59.5 岁,女性占 61.8%),通过对血压值和膳食摄入量做降秩回归来得到血压相关的膳食模式。因变量为 SBP[变量名 sbp,平均值为 136.6 mmHg(1 mmHg=0.133 kPa)]和 DBP(变量名 dbp,平均值为 78.3 mmHg),自变量为 10 个食物组的摄入量(大米、面食、杂粮、肉类及制品、家禽及制品、鱼类/海产品、蛋类及制品、新鲜蔬菜、豆制品、新鲜水果)。Stata 语句如下:

```
rrr sbp dbp, x (rice wheat other_staple meat poultry fish eggs fresh_veg soya fresh_fruit) loadings
```

Stata 输出结果主要分为三部分:①变量数汇总和一般线性回归结果;②主成分分析结果;③提取的降秩回归因子分别对因变量和自变量的方差解释百分比(图 1)。从结果可得,提取的两个降秩回归因子(膳食模式)共解释因变量(SBP 和 DBP) 0.9% 的变异,第一个因子和第二个因子分别解释因变量 0.5% 和 0.4% 的变异。第一个因子解释自变量组 10.3% 的变异,其中与肉类及制品和新鲜水果等食物组相关性相对较强;第二个因子解释自变量组 10.1% 的变异,其中与面食和豆制品等食物组

表 1 rrr 程序的语句选项及功能

语句	功能介绍	必要语句
depvars	因变量名列表:给定因变量组的信息(因变量必须多于 1 个),变量名之间用空格隔开	必要
indepvars	自变量名列表:给定自变量组的信息(自变量必须多于 1 个),变量名之间用空格隔开	必要
if	筛选条件:if 后面跟表达式,对纳入降秩回归分析的样本进行筛选或限制,例如 if age>50 表示只纳入年龄>50 岁的研究对象;默认值为纳入全样本	非必要
in	筛选条件:in 后面跟样本数据行数范围,例如 in 1/100 表示只纳入前 100 行样本数据;默认值为纳入全样本	非必要
rank()	保留的秩次数:设定最终提取的降秩回归因子的个数,括号内必须是>0 且不超过因变量个数的整数;默认值为因变量的个数	非必要
nostd	变量标准化(即 centering + scaling):加上 nostd 语句表示在运行 rrr 程序时不用对自变量和因变量进行标准化;默认值为不加 nostd,即在降秩回归前自动对每个自变量和因变量进行标准化	非必要
savevar	保留中间步骤生成的变量:加上 savevar 语句后, rrr 程序运行过程中生成的标准化后的自变量和因变量以及因变量的普通最小二乘估计回归预测值将被保存为新变量并入原有数据集;默认值为不加 savevar,即不保存这些中间变量	非必要
loadings	输出因子载荷:加上 loadings 语句后,提取出的一个或几个降秩回归因子的因子载荷将在输出界面分别展示;默认值为不加 loadings,即不输出因子载荷	非必要

表 2 rrr 程序的输出变量

变量名	变量介绍	必要输出
f1, f2, ...	降秩回归因子,例如 f1 表示提取的第一个降秩回归因子(即第一个主成分得分),该类变量的个数等于 rank 语句中设定的秩次数	必要
indepvar_std	标准化后的自变量,仅在不加 nostd 语句并且加 savevar 语句的情况下被保存	非必要
depvar_std	标准化后的因变量,仅在不加 nostd 语句并且加 savevar 语句的情况下被保存	非必要
depvar_hat	因变量的普通最小二乘估计回归预测值,仅在加 savevar 语句的情况下被保存	非必要

Variations of response variables explained by each Reduced Rank Regression factor

No	sbp	dbp	Average	Cumulative_average
f1	0.009889	0.000128	0.00500851	0.00500851
f2	0.000104	0.008005	0.00405443	0.00906295

Variations of predictors explained by each Reduced Rank Regression factor

No	rice	wheat	other_staple	meat	poultry	fish	eggs
f1	0.060974	0.182988	0.005874	0.248331	0.066804	0.027957	0.090982
f2	0.037589	0.544235	0.006963	0.020784	0.019039	0.143242	0.000962

	fresh_veg	soya	fresh_fruit	Average	Cumulative_average
f1	0.006196	0.051433	0.29253	0.10340696	0.10340696
f2	0.025417	0.211068	0.000645	0.10099438	0.20440134

Factor loadings of each Reduced Rank Regression factor

Factor loadings of f1

e(b)[1,10]

	rice_std	wheat_std	other_staple_std	meat_std	poultry_std	fish_std	eggs_std
y1	0.00507409	-0.0474064	-0.00306018	-0.06098048	-0.01637494	0.02607831	-0.01561847

	fresh_veg_std	soya_std	fresh_fruit_std
y1	0.0089115	0.03620031	-0.05013882

Factor loadings of f2

e(b)[1,10]

	rice_std	wheat_std	other_staple_std	meat_std	poultry_std	fish_std	eggs_std
y1	0.01968764	0.08195106	-0.00104537	-0.00941824	0.0093684	0.03652295	-0.01989887

	fresh_veg_std	soya_std	fresh_fruit_std
y1	-0.00181699	0.03557058	-0.01391092

图1 rrr程序输出结果:方差解释百分比和因子载荷

相关性相对较强。此分析所得的两个降秩回归因子对因变量的方差解释百分比偏低,可能与纳入的食物组较少有关(作为示例分析仅考虑了主要食物组的摄入量)。

此外,示范语句中加入了 loadings 语句以获得每个食物组的因子载荷(图1)。程序在原数据集里添加的两个降秩回归因子(膳食模式)得分变量 f1 和 f2(表3),就是通过各食物组的标准化后摄入量与其因子载荷乘积之和计算所得。rrr 程序提取的膳食模式因子载荷可被用于计算其他相似的膳食数据中相应的膳食模式得分,例如:

膳食模式 1 的计算公式: $f1 = \text{rice_std} \times (0.005\ 07) + \text{wheat_std} \times (-0.047\ 41) + \text{other_staple_std} \times (-0.003\ 06) + \text{meat_std} \times (-0.060\ 98) + \text{poultry_std} \times (-0.016\ 37) + \text{fish_std} \times (0.026\ 08) + \text{eggs_std} \times (-0.015\ 62) + \text{fresh_veg_std} \times (0.008\ 91) + \text{soya_std} \times (0.036\ 20) + \text{fresh_fruit_std} \times (-0.050\ 14)$ 。

四、讨论

本文介绍了降秩回归方法的基本原理,开发了

表3 rrr程序输出结果:变量汇总

序号	变量名	变量描述
1	sbp	因变量(SBP, mmHg)
2	dbp	因变量(DBP, mmHg)
3	rice	自变量(大米每周摄入量, g)
4	wheat	自变量(面食每周摄入量, g)
5	other_staple	自变量(杂粮每周摄入量, g)
6	meat	自变量(肉类及制品每周摄入量, g)
7	poultry	自变量(家禽及制品每周摄入量, g)
8	fish	自变量(鱼类/海产品每周摄入量, g)
9	eggs	自变量(蛋类及制品每周摄入量, g)
10	fresh_veg	自变量(新鲜蔬菜每周摄入量, g)
11	soya	自变量(豆制品每周摄入量, g)
12	fresh_fruit	自变量(新鲜水果每周摄入量, g)
13	f1	降秩回归因子 1
14	f2	降秩回归因子 2

基于 Stata 的 rrr 程序包,并通过中国慢性病前瞻性研究数据进行了膳食模式的示例分析。降秩回归目前主要应用于营养流行病学^[17],来研究从食物摄入量到营养素摄入量再到疾病结局的关联通路,或者建立食物摄入量-生物标记物-疾病的关联通路。

但是此方法同样适用于其他多变量关联通路的研究方向。

除降秩回归外,还有其他两种常用的多变量降维回归方法:主成分回归和偏最小二乘回归。主成分回归注重于对自变量组变异的解释,用自变量组的主成分得分和因变量组做回归;而偏最小二乘回归则兼顾两者的方差解释,用自变量组的主成分和因变量组的主成分做回归^[18]。相较而言,降秩回归更注重于对因变量组变异的解释,其因子得分对因变量的方差解释百分比往往优于另两种方法^[3]。值得注意的是,在实际应用中降秩回归秩次数的选择存在一定的主观性。保留的降秩回归因子越多(例如 $m=q$),其对因变量组变异的解释越好(即拟合度越高);保留的降秩回归因子越少,则降维的作用越明显,例如用一个降秩回归因子 f_1 去代表两个因变量(SBP 和 DBP)的信息。在实际操作中,使用者可以先采用程序默认的满秩设置(秩次等于因变量个数),根据主成分分析的结果并结合实际研究需要来选择想要保留的降秩回归因子的个数。例如考虑保留特征值 >1 的因子,或者当从第 i 个因子开始特征值大幅下降(即第 $i+1$ 个因子的特征值明显低于第 i 个因子),则考虑保留前 i 个因子。

本研究开发了首个采用降秩回归方法分析的 Stata 程序包,并提交到了 Stata 程序发布平台 SSC,供全球用户使用。本程序准确的实现了 SAS 软件的降秩回归分析功能。在本文的示例分析中,通过将 Stata 软件输出结果与 SAS 软件的 PLS(RRR)过程输出结果做横向对比,发现两个软件提取的降秩回归因子对自变量和因变量的方差解释百分比及因子载荷(SAS 软件中称为 Model Effect Weights)完全一致,而因子得分相差常数倍(f_1 和 f_2 分别相差 0.105 倍和 0.102 倍)。由于降秩回归因子得分没有量纲且其绝对数值没有实际意义,因此是完全等价的结果。

本程序开发的意义首先在于实现了 Stata 软件用户直接使用该软件做降秩回归的需求。目前 Stata 是国内外较为常用的统计软件之一,用户群体庞大,并且 Stata 软件编程语言与 SAS、R 软件等存在较大差异。在本程序开发前,大多数 Stata 软件用户只能通过其他软件进行降秩回归分析,因此本程序的开发打破了这一分析工具上的限制。其次,相较于已有的 SAS 程序和 R 程序,本程序提供了更为简便、灵活的语句设置。rrr 程序只有两个必要输入语句:自变量组和因变量组,因此主体语句结

构简洁易懂。此程序不仅遵循 Stata 官方程序格式,例如可以与 if、in 等语句兼容,还根据常见分析流程开发扩展选项,例如变量标准化功能。rrr 程序默认对变量进行标准化。如果数据已经经过标准化,或各变量的量纲一致,则可在命令中加 nostd 语句跳过该过程。第三,此程序的输出结果相比于其他已有程序更具系统性和结构性,有助于使用者更好地理解降秩回归的原理及分析步骤(即 OLS 估计、主成分分析和方差解释百分比计算)。此程序的 loadings 语句可以指定在运行完 rrr 程序后输出各个降秩回归因子的因子载荷,构建降秩回归因子的计算公式。该公式可直接应用到其他有相同自变量的数据库来生成降秩回归因子,从而增加研究间的可比性。

综上所述,降秩回归在公共卫生领域有着广泛的应用前景,而 Stata 程序的开发弥补了该软件用户无法直接使用该方法进行数据分析的不足,有助于其进一步推广使用。

利益冲突 所有作者声明无利益冲突

作者贡献声明 郑棒:软件开发、分析解释数据、起草文章、统计分析;刘琪:统计分析、对文章的知识性内容作批评性审阅;吕筠:对文章的知识性内容作批评性审阅、经费支持;余灿清:研究制订、软件开发、分析解释数据、起草文章、经费支持

参 考 文 献

- [1] Izenman AJ. Reduced-rank regression for the multivariate linear model[J]. J Multivar Anal, 1975, 5(2):248-264. DOI: 10.1016/0047-259X(75)90042-1.
- [2] van der Merwe A, Zidek JV. Multivariate regression analysis and canonical variates[J]. Can J Stat, 1980, 8(1): 27-39. DOI:10.2307/3314667.
- [3] Hoffmann K, Schulze MB, Schienkiewitz A, et al. Application of a new statistical method to derive dietary patterns in nutritional epidemiology[J]. Am J Epidemiol, 2004, 159(10):935-944. DOI:10.1093/aje/kwh134.
- [4] Shakya PR, Melaku YA, Page A, et al. Association between dietary patterns and adult depression symptoms based on principal component analysis, reduced-rank regression and partial least-squares[J]. Clin Nutr, 2020, 39(9):2811-2823. DOI:10.1016/j.clnu.2019.12.011.
- [5] Frank LK, Jannasch F, Kröger J, et al. A dietary pattern derived by reduced rank regression is associated with type 2 diabetes in an urban Ghanaian population[J]. Nutrients, 2015, 7(7):5497-5514. DOI:10.3390/nu7075233.
- [6] 殷召雪,任泽萍,徐小刚,等.基于降秩回归的血压相关膳食模式与老年人认知受损的关系[J].中华流行病学杂志, 2018, 39(6): 781-785. DOI: 10.3760/cma. j. issn. 0254-6450.2018.06.017.
Yin ZX, Ren ZP, Xu XG, et al. Association between blood pressure related dietary patterns and identified cognitive performance in the elderly Chinese—a study by reduced rank regression method[J]. Chin J Epidemiol, 2018, 39(6): 781-785. DOI:10.3760/cma.j.issn.0254-6450.2018.06.017.

[7] 程茅伟, 王惠君, 王志宏, 等. 基于降秩回归的膳食模式与代谢综合征相关性研究[J]. 营养学报, 2017, 39(2): 121-126. DOI:10.3969/j.issn.0512-7955.2017.02.004. Cheng MW, Wang HJ, Wang ZH, et al. Study on the correlation between dietary patterns and metabolic syndrome based on reduced rank regression analysis[J]. Acta Nutrimenta Sinica, 2017, 39(2): 121-126. DOI: 10.3969/j.issn.0512-7955.2017.02.004.

[8] Anderson TW, Rubin H. Estimation of the parameters of a single equation in a complete system of stochastic equations[J]. Ann Math Stat, 1949, 20(1): 46-63. DOI: 10.1214/aoms/1177730090.

[9] Anderson TW. Estimating linear restrictions on regression coefficients for multivariate normal distributions[J]. Ann Math Stat, 1951, 22(3):327-351. DOI: 10.1214/aoms/1177729580.

[10] SAS Institute Inc. SAS/STAT® 13.1 User's Guide[EB/OL]. Cary, NC: SAS Institute Inc., 2013. <https://support.sas.com/documentation/cdl/en/statug/66859/HTML/default/viewer.htm#titlepage.htm>.

[11] Addy C. rrr: reduced-rank regression[CP]. (2016) [2021-01-31]. <https://CRAN.R-project.org/package=rrr>.

[12] Chen K, Wang WJ, Yan J. rrpck: reduced-rank regression [EB/OL]. (2019-11-18). <https://cran.r-project.org/package=rrpck>.

[13] Mander A. PLSAS: Stata module to execute SAS partial least squares procedure (Windows only) [CP/OL]. Statistical Software Components S456810, Boston College Department of Economics, 2007(2012-06-21). <https://ideas.repec.org/c/boc/bocode/s456810.html>.

[14] Zheng B, Yu CQ. RRR: Stata module to perform Reduced rank regression[CP/OL]. Statistical Software Components S458882, Boston College Department of Economics, 2020. <https://ideas.repec.org/c/boc/bocode/s458882.html>.

[15] Kohler U, Luniak M. Data Inspection using Biplots[J]. Stata J, 2005, 5(2):208-223. DOI:10.1177/1536867X0500500206.

[16] 李立明, 吕筠, 郭彧, 等. 中国慢性病前瞻性研究:研究方法和调查对象的基线特征[J]. 中华流行病学杂志, 2012, 33(3): 249-255. DOI: 10.3760/cma.j.issn.0254-6450.2012.03.001. Li LM, Lv J, Guo Y, et al. The China Kadoorie Biobank: related methodology and baseline characteristics of the participants[J]. Chin J Epidemiol, 2012, 33(3): 249-255. DOI:10.3760/cma.j.issn.0254-6450.2012.03.001.

[17] 程茅伟, 张兵. 应用降秩回归评估膳食模式与代谢紊乱性疾病相关性的研究进展[J]. 营养学报, 2020, 42(1):83-91. DOI:10.3969/j.issn.0512-7955.2020.01.017. Cheng MW, Zhang B. Application of reduced rank regression in evaluating the association between dietary patterns and metabolic diseases[J]. Acta Nutrimenta Sinica, 2020, 42(1): 83-91. DOI: 10.3969/j.issn.0512-7955.2020.01.017.

[18] 罗晓, 陈耀, 孙优贤. 基于统计回归的质量推断方法[J]. 信息与控制, 2001, 30(5):422-426. DOI:10.3969/j.issn.1002-0411.2001.05.008. Luo X, Chen Y, Sun YX. Statistical-regression-based quality inference methods[J]. Inform Control, 2001, 30(5): 422-426. DOI:10.3969/j.issn.1002-0411.2001.05.008.

中华预防医学会流行病学分会第八届委员会组成人员名单

(按姓氏笔画排序)

顾问	刘天锡	汪华	陆林	姜庆五	贺雄				
名誉主任委员	李立明								
主任委员	詹思延								
副主任委员	叶冬青	冯子健	何纳	何耀	沈洪兵	胡永华			
常务委员	王岚	王子军	王全意	王素萍	代敏	吕筠	朱凤才	江宇	
	许国章	李立明	李亚斐	杨晓明	杨维中	吴凡	吴先萍	汪宁	
	张建中	陈坤	赵根明	胡志斌	段广才	俞敏	施小明	唐金陵	
	曹务春	谭红专							
委员	丁淑军	么鸿雁	王蓓	王建明	毛琛	仇小强	方向华	田文静	
	白亚娜	吕繁	庄贵华	刘玮	刘运喜	刘雅文	刘殿武	许汴利	
	孙业桓	苏虹	李琦	李文庆	李石柱	李佳圆	杨西林	杨敬源	
	吴尊友	吴寰宇	邱洪斌	余宏杰	张本	张军	张卫东	张毓洪	
	陈可欣	陈维清	邵中军	欧剑鸣	周宝森	官旭华	孟蕾	项永兵	
	赵亚双	胡东生	施榕	姜勇	姜晶	袁萍	贾存显	贾崇奇	
	高立冬	郭卫东	郭秀花	曹广文	梁娴	寇长贵	彭霞	韩秀敏	
	程锦泉	程慧健	曾小云	雷立健	蔡建芳	缪小平	潘安	戴江红	
	魏文强								
秘书长	王岚								
秘书	余灿清	李银鸽							