

广义估计方程与混合线性模型在Python中的实现



焦奎壮, 马煦晰, 马小茜, 刘朝屹, 张青, 马露

武汉大学公共卫生学院 (武汉 430071)

【摘要】目的 探索纵向数据分析中广义估计方程和混合线性模型在 Python 中的实现方法, 拓展 Python 软件在统计分析中的应用。**方法** 通过 Python 软件构建广义估计方程和混合线性模型, 以环境流行病学的研究实例探索大气颗粒物 $PM_{2.5}$ 对肺功能的影响, 并与 R 软件分析结果进行比较。**结果** $PM_{2.5}$ 是肺功能降低的危险因素, $PM_{2.5}$ 每升高 $1 \mu g/m^3$, 研究对象 2 天后第 1 秒用力呼气容积减少 8 mL。Python 软件可用 statsmodels 库进行广义估计方程和混合线性模型的分析, 实现过程程序语言简洁。与 R 软件相比, 程序逻辑具有一定的相似性, 参数估计和置信区间的计算结果基本一致, Python 结果可信。**结论** Python 软件可灵活构建广义估计方程和混合线性模型, 在实际研究中有一定参考价值。

【关键词】 Python; 广义估计方程; 混合线性模型; 纵向数据

Implementation of generalized estimating equations and mixed linear models in Python

Kui-Zhuang JIAO, Xu-Xi MA, Xiao-Qian MA, Chao-Yi LIU, Qing ZHANG, Lu MA

School of Public Health, Wuhan University, Wuhan 430071, China

Corresponding Author: Lu MA, Email: malu@whu.edu.cn

【Abstract】Objective Explore the implementation of generalized estimation equations (GEE) and mixed linear models (MLM) in longitudinal data analysis using Python software, and expand its application in statistical analysis. **Methods** GEE and MLM were constructed by Python software to explore the impact of $PM_{2.5}$ on lung function (forced expiratory volume in 1 second, FEV1) with an example of environmental epidemiology, and compared with the results of R software. **Results** With $PM_{2.5}$ increases of $1 \mu g/m^3$, the FEV1 of the subjects decreased by 8 mL after 2 days. Python software can use a statsmodels library to analyze MLM and GEE, and the program language is concise, the program logic has a certain similarity when compared with R, the calculation results of parameter estimation and confidence interval are almost the same, and the Python result is reliable. **Conclusion** Python software can flexibly construct MLM and GEE, which has a certain reference value in practical research.

【Keywords】 Python; Generalized estimation equation; Mixed linear model; Longitudinal data

DOI: [10.12173/j.issn.1004-5511.202203007](https://doi.org/10.12173/j.issn.1004-5511.202203007)

基金项目: 湖北省卫生健康委 2021—2022 年度科研项目 (WJ2021F103)

通信作者: 马露, 博士, 副教授, 硕士研究生导师, Email: malu@whu.edu.cn

纵向数据是流行病学研究中最常见的资料类型之一, 常见于队列研究、定群研究等研究设计中。这些研究设计中暴露因素或健康结局指标均具有随时间变化而变化的特点, 为探索暴露因素与健康结局间的关系, 往往需对研究对象进行随访或重复测量。针对同一研究对象的多次测量, 研究结果间通常存在相关关系, 若不满足独立性的条件, 不适用一般线性模型或广义线性模型, 且当观测值存在缺失时, 重复测量方差分析也不适用。因此, 广义估计方程 (generalized estimating equations, GEE) 和混合线性模型 (mixed linear model, MLM) 被广泛应用于纵向数据的统计分析。目前, 主流的统计分析软件如 SAS、SPSS 和 R 等均能实现 GEE 和 MLM 的建模分析^[1-3]。Python 作为一款开源免费软件, 因其强大的大数据处理第三方库 (Pandas、Numpy、Scipy 等)、内存优化系统和丰富的应用场景 (爬虫等), 可方便快速地实现数据的获取、清洗、管理和分析, 显著缩短数据分析时间^[4-5], 近年来越来越受到国内科研工作者的欢迎。随着我国医疗系统信息化建设的快速推进, 医疗大数据的智能化统计分析是必然的发展趋势^[6]。运用 Python 软件实现流行病学研究中的统计分析, 目前尚不多见。本研究以 Python 3.8.5 中的 statsmodels 库为例, 通过研究实例介绍 GEE 和 MLM 在 Python 软件中的实现方法。同时, 采用 R 4.0.5 软件中的 geepack 包和 lmerTest 包构建 GEE 与 MLM 模型^[7], 作为本次 Python 结果的对照, 验证 Python 输出结果是否正确。

1 资料与方法

1.1 资料来源

为研究某地区大气颗粒物 PM_{2.5} 对肺功能的影响, 收集该地区连续十日的日均 PM_{2.5} 浓度、温湿度和研究对象的肺功能数据。其中, PM_{2.5} 浓度、温湿度数据分别来自武汉市生态环境保护局 (<http://hbj.wuhan.gov.cn/>) 和湖北省气象局 (<http://hb.cma.gov.cn/>), 均采用 Python 爬虫收集。

采用方便抽样方法抽取研究对象, 使用第 1 秒用力呼气容积 (forced expiratory volume in one second, FEV1) 作为肺功能评估指标, 使用肺功能测试仪于每日 7:30 AM 由研究对象自测, 测量 10 日后通过软件将数据读入计算机。以某位研究对象 (20 岁) 为例, 显示十天连续测量数据, 详见表 1。

1.2 模型构建

本研究通过控制研究对象年龄、BMI 和大气温湿度, 评价 PM_{2.5} 对研究对象肺功能的影响, 以 PM_{2.5} 单污染物滞后 2 天的暴露模型为例展示 GEE 与 MLM 在 Python 软件中的实现。

1.2.1 GEE 模型构建

GEE 模型是 Liang 和 Zeger 于 1986 年在拟似然的基础上对广义线性模型的推广, 旨在分析纵向数据中因素对总体平均水平的影响^[8-10]。本研究以 FEV1 为因变量, 记为 Y_{ij}, 表示有 i 个研究对象 (1, ..., n), 每个研究对象有 j 个观察值 (1, ..., p), 协变量记为 X_{ij}。构建如下模型:

$$(1) E(Y_{ij}) = \mu_{ij}, X'_{ij}\beta = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \dots + \beta_j X_{ijp}, g(\mu_{ij}) \text{ 为链接函数。}$$

表1 连续10日颗粒物、气象及研究对象数据资料

Table 1. Data of particulate matter, meteorology and research objects in ten consecutive days

时间	BMI (kg/m ²)	温度 (°C)	湿度 (%rh)	PM _{2.5} (µg/m ³)	FEV1 (L)
第1天	16.71	21.92	61.75	8.17	3.25
第2天	16.71	22.32	66.64	11.83	3.01
第3天	16.71	23.51	70.17	19.09	3.43
第4天	16.71	22.62	72.16	39.11	3.14
第5天	16.71	22.21	73.45	21.05	3.14
第6天	16.71	22.83	72.29	38.50	3.08
第7天	16.71	23.85	65.12	27.38	3.26
第8天	16.71	23.01	61.88	34.00	2.86
第9天	16.71	21.88	61.65	25.58	3.38
第10天	16.71	21.61	62.07	20.67	3.06

注: BMI: 体质量指数; FEV1: 第1秒用力呼气容积 (forced expiratory volume in one second)

建立 Y_{ij} 与 X_{ij} 间的函数关系，链接函数的选择主要有以下几种形式：①因变量服从正态分布，链接函数选择 identify；②因变量服从二项分布，链接函数选择 logit；③因变量服从泊松分布，链接函数选择 log；④因变量服从负二项分布，链接函数选择 negativebinomial。

(2) $\text{Var}(Y_{ij}) = v(\mu_{ij}) \cdot \phi$ ，建立 Y_{ij} 的方差与均值间的函数关系。

(3) $R_i(\alpha)$ 是 Y_{ij} 的作业相关矩阵，表示因变量各次的重复测量值间的相关性大小，作业相关矩阵包括以下几种形式：①可交换，又称等相关，即任意两次不同时间的观测值间相关是相等的；②相邻相关，即只有相邻时间的两次测量值之间具有相关性；③自相关，即相关与间隔次数有关，相隔次数越长，相关关系越小；④不确定型相关，即相关矩阵非对角线上的元素均不等；⑤独立，即因变量之间不相关。

根据 Liang 和 Zeger 的定义，构建如下 GEE 模型：

$$S(\beta; \alpha, \phi) = \sum_i \left(\frac{\partial \mu_i}{\partial \beta} \right) V_i^{-1}(\mu_i; \alpha) (y_i - \mu_i) = 0$$

V_i 是 Y_i 的协方差矩阵， $V_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2}$ 。根据给定的 α 和 ϕ 的估计值，通过迭代重复加权，采用最小二乘法，求解上述方程，最后得出 β 的一致性估计。

1.2.2 MLM 模型构建

MLM 是一般线性模型 ($Y = X\beta + \varepsilon$) 的拓展^[11-12]，其保留了传统线性模型中的残差需满足正态性的

假定，而对独立性与方差齐性不做要求^[13]，引入随机效应部分 $Z\gamma$ ，可表达为： $Y = X\beta + Z\gamma + \varepsilon$ ， Y 表示因变量测量值的向量， X 为固定效应设计矩阵， β 为固定效应参数向量， Z 为随机效应设计矩阵， γ 为随机效应参数向量，服从均值向量为 0、方差 / 协方差矩阵为 G 的正态分布，表示为 $\gamma \sim N(0, G)$ ，随机效应主要有以下三种^[14-15]：①随机回归系数带来的随机效应；②随机截距带来的随机效应；③随机回归系数和随机截距带来的随机效应。 ε 为随机误差向量，服从均值向量为 0、方差 / 协方差矩阵为 R 的正态分布，即 $\varepsilon \sim N(0, R)$ 。

1.3 模型验证

采用 R 4.0.5 软件的 geepack 包、lmerTest 包分别构建 GEE 与 MLM 模型，作为参照，验证 Python 输出结果是否正确。

2 结果

2.1 GEE 建模及参数估计结果

本研究因变量 FEV1 为定量连续数据，服从正态分布，故链接函数选择 identify，根据以上定义，可构建如下 GEE 模型：

$$FEV1 = \beta_0 + \beta_1 PM_{2.5} + \beta_2 Age + \beta_3 BMI + \beta_4 Temperature + \beta_5 Humidity + CORR + \varepsilon$$

β_0 为截距， β_1 、 β_2 、 β_3 、 β_4 、 β_5 为各协变量的回归系数，CORR 表示作业相关矩阵， ε 表示残差项，相关代码见框 1，GEE 参数估计结果如表 2 所示。

Python	R
import pandas as pd	library(geepack)
import statsmodels.api as sm	dt1 <- read.csv('E:base_data.csv', encoding='UTF-8',
import statsmodels.formula.api as smf	stringsAsFactors=F)
data = pd.read_csv('E:base_data.csv')	fit <- geeglm(FEV1~Age+BMI+PM2_5+Temperature +
fam = sm.families.Gaussian()	Humidity,id=ID,data=dt1,corstr=
ind = sm.cov_struct.Exchangeable()	'exchangeable',family='gaussian')
mod = smf.gee("FEV1~BMI + Temperature + Humidity + Age +	cc <- coef(summary(fit))
PM2_5", "ID", data=data, cov_struct=ind,family=fam)	citab <- with(as.data.frame(cc), cbind(lwr=Estimate*1.96*
res = mod.fit()	Std.err, upr=Estimate+1.96*Std.err))
print(res.summary())	rownames(citab) <- rownames(cc)

框1 广义估计方程代码呈现
Box 1. GEE codes in Python and R

表2 广义估计方程参数估计结果

Table 2. Results of GEE parameter estimation

参数	Python			R		
	估计值 (95%CI)	z值	P值	估计值 (95%CI)	χ^2 值	P值
截距	2.741 (-4.067, 9.549)	0.79	0.430	2.748 (-4.063, 9.558)	0.63	0.429
PM _{2.5}	-0.008 (-0.014, -0.001)	-2.31	0.021	-0.008 (-0.014, -0.001)	5.35	0.021
年龄	0.033 (-0.340, 0.405)	0.17	0.863	0.033 (-0.340, 0.405)	0.03	0.864
BMI	-0.008 (-0.113, 0.096)	-0.16	0.875	-0.008 (-0.113, 0.096)	0.03	0.875
温度	-0.012 (-0.055, 0.031)	-0.56	0.576	-0.012 (-0.056, 0.031)	0.32	0.569
湿度	0.009 (-0.002, 0.019)	1.61	0.107	0.009 (-0.002, 0.019)	2.62	0.106

2.2 MLM建模及参数估计结果

根据 MLM 定义, 可构建如下 MLM 模型:

$$FEV1 = \beta_0 + \beta_1 PM_{2.5} + \beta_2 Age + \beta_3 BMI + \beta_4 Temperature + \beta_5 Humidity + Z\gamma + \varepsilon$$

β_0 为总截距, β_1 、 β_2 、 β_3 、 β_4 、 β_5 为各协变量的回归系数, ε 为残差项, $Z\gamma$ 是随机效应, 对应研究对象的个体差异带来的随机截距。相关代码见框 2, MLM 参数估计结果如表 3 所示。

根据模型中自变量的检验结果可知, PM_{2.5} 是肺功能降低的危险因素, PM_{2.5} 每升高 1 $\mu\text{g}/\text{m}^3$, 研究对象 2 天后的 FEV1 减少 8 mL ($P < 0.05$)。GEE 模型中, Python 与 R 输出的检验统计量分别为 z 和 χ^2 , 同一自变量参数检验结果与 P 值基本一致; MLM 模型中, Python 与 R 输出的检验统计量分别为 z 和 t , 同一自变量参数检验结果与 P 值基本一致, 表示 Python 输出结果基本可信。

Python	R
import pandas as pd	library(lmerTest)
import statsmodels.api as sm	dt1 <- read.csv('E:base_data.csv', encoding = 'UTF-8', stringsAsFactors = F)
import statsmodels.formula.api as smf	fit1 <- lmer(FEV1 ~ Age+BMI+PM2_5+Temperature + Humidity+(1 ID), data=dt1)
data = pd.read_csv('E:base_data.csv')	summary(fit1)
mod_lme = smf.mixedlm("FEV1 ~ BMI + Temperature + Humidity + Age + PM2_5", groups=data["ID"], data=data)	
modf_lme = mod_lme.fit()	
print(modf_lme.summary())	

框2 混合线性模型代码呈现

Box 2. MLM codes in Python and R

表3 混合线性模型参数估计结果

Table 3. Results of MLM parameter estimation

参数	Python			R		
	估计值 (95%CI)	z值	P值	估计值 (95%CI)	t值	P值
截距	2.746 (-4.126, 9.617)	0.78	0.430	2.746 (-4.126, 9.617)	0.78	0.450
PM _{2.5}	-0.008 (-0.013, -0.002)	-2.89	0.004	-0.008 (-0.013, -0.002)	-2.89	0.005
年龄	0.033 (-0.319, 0.384)	0.18	0.855	0.033 (-0.318, 0.384)	0.18	0.859
BMI	-0.008 (-0.107, 0.091)	-0.17	0.868	-0.008 (-0.107, 0.091)	-0.17	0.872
温度	-0.012 (-0.091, 0.066)	-0.31	0.756	-0.012 (-0.091, 0.066)	-0.31	0.757
湿度	0.009 (-0.003, 0.020)	1.49	0.137	0.009 (-0.003, 0.020)	1.49	0.140

3 讨论

本文结合流行病学中的研究实例, 简要介绍了 GEE 和 MLM 在 Python 中的具体操作, 拓展了纵向数据分析的实现方法。尽管纵向数据在 Python 与 R 软件中的实现代码有些许不同, 但输出的参数和 P 值检验结果基本一致, 反映 Python 输出结果可信。在不同的软件中, 相同的方法模型输出的统计量有所差异, 这是开发人员的统计检验倾向导致的。鉴于 Z 统计量值的平方等于 Wald χ^2 统计量的值, 因此 Wald χ^2 检验是等价于 Z 检验的^[16]; 在大样本 ($n > 50$) 的情况下, t 检验与 Z 检验结果也是近似的^[17], 所以软件输出参数统计量的不同不影响参数检验结果。

Python 在数据分析方面存在诸多优点, 其完善的作图功能以及丰富的数据分析库、机器学习库的发展, 越来越符合大数据背景下的数据分析要求。例如, 陈伟等使用 Python 相似度分析与标签云分析方法进行文本数据分析, 拓展了大数据审计的研究方向^[18]; 杨俊秀等采用 Python 的 matplotlib 库、numpy 库和 scikit-learn 库完成了数据的可视化、整理与分析, 显著提高了高频电子线路实验数据的处理分析效率^[5]。但 Python 也存在着不足之处, 其用作数据分析工具的时间较短, 更多的数据分析方法程序有待进一步开发及完善。本研究使用的 statsmodels 库目前可实现简单线性模型、广义线性模型、GEE、MLM 等大部分模型的建模。随着 statsmodels 库的进一步开发, 相信未来 statsmodels 库可满足更多的建模需求。

综上, 将 Python 用于环境流行病学, 在实现数据获取、处理与分析的过程中, 统一了语言环境, 避免了数据在不同平台间的转换, 提高了数据分析的效率。使用 Python 软件可灵活实现 GEE 和 MLM 的统计建模, 在实际研究中有一定参考价值。

参考文献

- 1 周婷, 兰蓝, 邱建青, 等. GEE、GLMM 和 MLM 分析卫生重复测量资料的效果比较[J]. 现代预防医学, 2017, 44(16): 2881–2885, 2899. [Zhou T, Lan L, Qiu JQ, et al. Effect comparison of GEE, GLMM, and MLM in analyzing repeated measures of health-related field[J]. Modern Preventive Medicine, 2017, 44(16): 2881–2885,

- 2899.] DOI: CNKI:SUN:XDYF.0.2017–16–001.
- 2 陈樱, 黄碧芬, 郑建清, 等. 基于 SAS 软件的混合效应模型实现重复测量数据的 Meta 分析[J]. 中国循证医学杂志, 2019, 19(8): 998–1005. [Chen Y, Huang BF, Zheng JQ, et al. Meta-analysis of repeated measurement data based on mixed effects model of SAS software[J]. Chinese Journal of Evidence-Based Medicine, 2019, 19(8): 998–1005.] DOI: 10.7507/1672–2531.201902004.
- 3 特维斯克. 实用流行病学纵向数据分析方法 (第 2 版)[M]. 北京: 人民卫生出版社, 2016. [Twisk. Applied longitudinal data analysis for epidemiology (Version 2)[M]. Beijing: People's Medical Publishing House, 2016.]
- 4 平凯珂, 陈平雁. Python 与 R 语言联合应用的实现[J]. 中国卫生统计, 2017, 34(2): 358–360. [Ping KK, Chen PY. Realization of joint application of Python and R language[J]. Chinese Journal of Health Statistics, 2017, 34(2): 358–360.] http://med.wanfangdata.com.cn/Paper/Detail?id=PeriodicalPaper_zgwstj201702054.
- 5 杨俊秀, 赵文来, 姚青, 等. 高频电子线路实验数据的 Python 处理分析[J]. 实验技术与管理, 2021, 38(10): 227–231, 240. [Yang JX, Zhao WL, Yao Q, et al. Experimental data processing and analysis of high frequency electronic circuit by Python[J]. Experimental Technology and Management, 2021, 38(10): 227–231, 240.] DOI: 10.16791/j.cnki.sjg.2021.10.041.
- 6 周江杰, 王胜锋, 李立明. Python 爬虫技术在信息流行病学中的应用[J]. 中华流行病学杂志, 2020, 41(6): 952–956. [Zhou JJ, Wang SF, Li LM. Application of Python web crawler technology in infodemiology[J]. Chinese Journal of Epidemiology, 2020, 41(6): 952–956.] DOI: 10.3760/cma.j.cn112338–20190901–00643.
- 7 朱雪宁. R 语言: 从数据思维到数据实战[M]. 北京: 中国人民大学出版社, 2018. [Zhu XN. Playing with R: data thinking to practice[M]. Beijing: China Renmin University Press, 2018.]
- 8 朱玉, 王静, 何倩. 广义估计方程在 SPSS 统计软件中的实现[J]. 中国卫生统计, 2011, 28(2): 199–201. [Zhu Y, Wang J, He Q. Implementation of generalized estimating equation in SPSS statistical software[J]. Chinese Journal of Health Statistics, 2011, 28(2): 199–201.] DOI: 10.3969/j.issn.1002–3674.2011.02.031.
- 9 Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation

- approach[J]. *Biometrics*, 1988, 44(4): 1049–1060. DOI: [10.2307/2531734](https://doi.org/10.2307/2531734).
- 10 李洪艳, 谭珊, 高晓, 等. 基于广义估计方程的婴儿超重的影响因素分析 [J]. *中国卫生统计*, 2016, 33(2): 222–225, 230. [Li HY, Tan S, Gao X, et al. A study about the influence factors of infant overweight based on generalized estimating equation[J]. *Chinese Journal of Health Statistics*, 2016, 33(2): 222–225, 230.] DOI: [CNKI:SUN:ZGWT.0.2016-02-011](https://doi.org/CNKI:SUN:ZGWT.0.2016-02-011).
- 11 陈博文, 赵培信, 唐新蓉, 等. 线性混合效应模型的有效稳健经验似然推断 [J]. *应用数学*, 2020, 33(4): 886–893. [Chen BW, Zhao PX, Tang XR, et al. Efficient and robust empirical likelihood inference for linear mixed effects models[J]. *Mathematica Applicata*, 2020, 33(4): 886–893.] DOI: [10.13642/j.cnki.42-1184/o1.2020.04.008](https://doi.org/10.13642/j.cnki.42-1184/o1.2020.04.008).
- 12 高莹, 沈玉, 周慧婵, 等. 个体 PM2.5 暴露与老年人肺功能关系 [J]. *环境卫生学杂志*, 2019, 9(5): 482–488. [Gao Y, Shen Y, Zhou HC, et al. Relationship between individual PM2.5 exposure and lung function in the elderly[J]. *Journal of Environmental Hygiene*, 2019, 9(5): 482–488.] DOI: [10.13421/j.cnki.hjwsxzz.2019.05.014](https://doi.org/10.13421/j.cnki.hjwsxzz.2019.05.014).
- 13 杨珉, 李晓松. 医学和公共卫生研究常用多水平统计模型 [M]. 北京: 北京大学医学出版社, 2007. [Yang M, Li XS. *Multilevel statistical models commonly used in medical and public health research*[M]. Beijing: Peking University Medical Press, 2007.]
- 14 庄东辰, 茆诗松. 混合系数线性模型的参数估计 [J]. *应用概率统计*, 1996, (1): 81–87. [Zhuang DC, Mao SS. Parameter estimation of mixed coefficient linear model[J]. *Chinese Journal of Applied Probability and Statistics*, 1996, (1): 81–87.] DOI: [CNKI:SUN:YYGN.0.1996-01-011](https://doi.org/CNKI:SUN:YYGN.0.1996-01-011).
- 15 余松林. 混合线性模型的应用 [J]. *中国医院统计*, 2006, 13(1): 70–75. [Yu SL. Application of mixed linear model[J]. *Chinese Journal of Hospital Statistics*, 2006, 13(1): 70–75.] DOI: [10.3969/j.issn.1006-5253.2006.01.030](https://doi.org/10.3969/j.issn.1006-5253.2006.01.030).
- 16 Wallis, Sean. z-squared: the origin and application of χ^2 [J]. *Journal of Quantitative Linguistics*, 2013, 20(4): 350–378. DOI: [10.1080/09296174.2013.830554](https://doi.org/10.1080/09296174.2013.830554).
- 17 智冬晓, 许晓娟, 张皓博. z 检验与 t 检验方法的比较 [J]. *统计与决策*, 2014, (20): 31–34. [Zhi DX, Xu XJ, Zhang HB. Comparison of Z-test and t-test[J]. *Statistics and Decision*, 2014, (20): 31–34.] DOI: [10.13546/j.cnki.tjyjc.2014.20.007](https://doi.org/10.13546/j.cnki.tjyjc.2014.20.007).
- 18 陈伟, 勾东升, 徐发亮. 基于文本数据分析的大数据审计方法研究 [J]. *中国注册会计师*, 2018, (11): 80–84. [Chen W, Gou DS, Xu FL. Research on big data audit method on basis of text data analysis[J]. *The Chinese Certified Public Accountant*, 2018, (11): 80–84.] DOI: [10.16292/j.cnki.issn1009-6345.2018.11.016](https://doi.org/10.16292/j.cnki.issn1009-6345.2018.11.016).

收稿日期: 2022 年 03 月 02 日 修回日期: 2022 年 03 月 17 日
本文编辑: 桂裕亮 黄 笛

引用本文: 焦奎壮, 马煦晰, 马小茜, 等. 广义估计方程与混合线性模型在 Python 中的实现 [J]. *医学新知*, 2022, 32(5): 333–338. DOI: [10.12173/j.issn.1004-5511.202203007](https://doi.org/10.12173/j.issn.1004-5511.202203007)
Jiao KZ, Ma XX, Ma XX, et al. Implementation of generalized estimating equations and mixed linear models in Python[J]. *Yixue Xinzhi Zazhi*, 2022, 32(5): 333–338. DOI: [10.12173/j.issn.1004-5511.202203007](https://doi.org/10.12173/j.issn.1004-5511.202203007)