

目录

附录 A AutoModelR: 操作指南与结果解释 1

1. 软件简介 2

 1.1 数据读取与预处理 2

 1.2 LASSO 自动变量选择 2

 1.3 中介/调节变量识别 2

 1.4 中介顺序与调节插入搜索 2

 1.5 拟合路径模型并输出可复现结果 2

2. 操作界面与输入参数 2

 2.1 界面说明: 2

3. 分析日志与解释 2

 3.1 下为一次完整运行日志的节选及其逐步解释: 2

4. 导出结果与可复现性 5

5. 可复现性说明 5

6. 回归模型整体指标解读 5

7. 系数估计及相关指标 6

8. 残差分析指标（拟合质量与诊断） 6

9. 举例说明 7

1. 软件简介

AutoModelR 是一款基于图形界面的心理建模工具，支持以 .xlsx 或 .csv 格式输入数据，自动完成以下流程：

1.1 数据读取与预处理

1.2 LASSO 自动变量选择

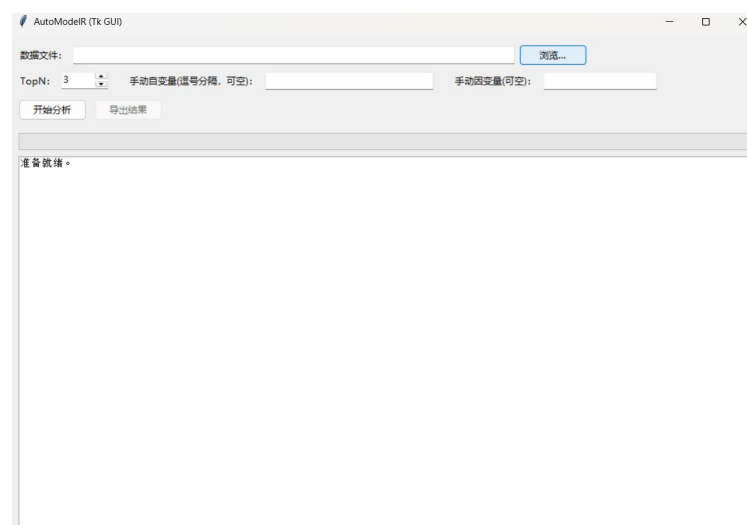
1.3 中介/调节变量识别

1.4 中介顺序与调节插入搜索

1.5 拟合路径模型并输出可复现结果

2. 操作界面与输入参数

软件主界面如下图所示：



2.1 界面说明：

2.1.1 数据文件：选择 Excel 或 CSV 数据文件；

2.1.2 TopN：用于控制 LASSO 返回的前 N 个变量组合；

2.1.3 手动自变量 / 因变量：可选输入，若填写将跳过自动变量搜索。

3. 分析日志与解释

3.1 下为一次完整运行日志的节选及其逐步解释：

3. 1. 1 读取数据与预处理

```
[1/6] 读取数据：E:/.../训练数据.xlsx
原始维度：(1047, 8)

[2/6] 预处理...
预处理后维度（仅数值列）：(1047, 8)
```

上图所示表示所有列均为数值变量，数据完整，无需筛选，如果使用的是量表数据请进行总分加和，并保障每列数据量相同；如果使用的是生物学数据请进行手动分组加和或加权。

括号中表示（被试数量/行数和变量数/列数）被试数量和变量数

尽量不要使用中文名称，尽量只选取关键变量（否则软件运行时间会很长）。

3. 1. 2 LASSO TopN 组合搜索

```
[3/6] LASSO TopN 组合搜索...
TopN 变量组合：
  y          x      r2
FNE          SA 0.502144
SA          FNE 0.501316
COG AB, IND, SSC, AIU 0.465062
```

上图所示结果表示：最佳模型 1 为 SA → FNE，决定系数 $R^2 \approx 0.502$ ，说明 SA 可解释 FNE 的约 50% 变异。同时对称方向（如 FNE → SA）也进行了建模对比。第 3 行为多个变量联合预测的结果。

决定系数越大越好，但要注意这里可能存在共线性问题，如 $M=M_1+M_2$ 是，输出最佳模型可能为 $M \sim M_1, M_2$ ，这种就存在共线性问题。

3. 1. 3 中介/调节判别

[4/6] 中介/调节判别...

var	type	p_mediation	p_moderation
AB	none	0.406185	0.669302
BB	none	0.859626	0.963834
IND	none	0.982244	0.707671
COG	moderator	0.997471	0.049588
SSC	none	0.892016	0.822601
AIU	none	0.890703	0.857322

在确定了最佳自变量和因变量后，中介效应判别基于间接路径 $a \times b$ 的显著性（例如 Sobel 检验）调节效应基于交互项的显著性（例如 $X \times W$ ），如上图所示，仅 COG 被识别为调节变量（ $p \approx 0.05$ ）。

type 表示该变量可能作为什么类型的（中介/调节变量），与自变量共同影响因变量。可以再后面对应 p 值中查看具体 p 值。

3.1.4 多中介与调节结构搜索

[5/6] 多中介顺序与调节位置搜索...

最佳中介顺序：(无)

调节插入：moderator COG on SA \rightarrow FNE； 调节变量：COG

上图中表示未识别出稳定的中介路径，但推断出 COG 在 SA \rightarrow FNE 路径上起调节作用。

这部分就是在确定了自变量和因变量的情况下，如何形成最佳变量模型，若为中介变量，则显示为 $X \sim M \sim Y$

3.1.5 拟合模型与输出摘要

Simple regression:

OLS Regression Results

```
=====
Dep. Variable:          FNE    R-squared:          0.503
Model:                 OLS    Adj. R-squared:       0.502
...
coef(SA) = 0.7092, p < 0.001
```

在上图结果中 $FNE = 0.709 \times SA + \varepsilon$ ，SA 显著预测 FNE，回归系数为 0.709，拟合度高。残差无自相关，偏态与峰度处于可接受范围。

4. 导出结果与可复现性

点击“导出结果”按钮后，以下文件将被保存：

文件名	内容说明
xy_top.csv	LASSO 选出的变量组合
roles.csv	每个变量的中介/调节判别结果
med_order.json	中介变量排列顺序
mod_pos.json	调节变量插入路径
fit.txt	回归模型摘要与参数估计

5. 可复现性说明

项目	要求
输入格式	.xlsx 或 .csv
操作系统	Windows 10/11
安装要求	无需安装 Python 或其他依赖
输出可读性	所有结果均以 UTF-8 文本或表格输出

6. 回归模型整体指标解读

项目	含义	作用	一般可接受标准
R-squared (R²)	决定系数，表示自变量能解释因变量变异的比例	衡量模型拟合优度	≥ 0.3 表示中等效果， ≥ 0.5 通常视为良好

项目	含义	作用	一般可接受标准
Adj. R-squared	调整后的 R^2 ，考虑了变量个数对 R^2 的影响	防止因变量过多导致虚高	通常略低于 R^2 ， > 0.5 表示模型稳健
F-statistic	回归方程的整体显著性检验值	用于检验模型是否整体显著	F 值越大越好；需结合 p 值判断
Prob (F-statistic)	F 值对应的显著性水平 (p 值)	判断整体模型是否成立	$p < 0.05$ 表示模型整体显著

7. 系数估计及相关指标

项目	含义	作用	可接受范围或判断
coef (回归系数)	每个自变量对因变量的线性影响程度	预测因变量的变化方向和幅度	与实际意义和方向一致为佳
std err (标准误差)	回归系数估计值的标准误差	衡量系数稳定性	越小越好 (说明估计更可靠)
t (t 值)	系数与其标准误差比值，用于检验系数是否显著	用于单个自变量显著性检验	t 值绝对值 > 1.96 ($p < 0.05$) 为显著
**P>	t	(p 值) **	自变量对应的显著性检验 p 值
[0.025, 0.975] (置信区间)	回归系数的 95% 置信区间	判断估计值可能波动范围	区间不跨 0，通常认为显著有效

8. 残差分析指标 (拟合质量与诊断)

指标	含义	作用	判断标准
Omnibus	联合检验残差是否	正态性检验之	$p > 0.05$ 表示残差正态

指标	含义	作用	判断标准
	正态分布	一	分布
Prob(Omnibus)	Omnibus 对应 p 值	越大越好	> 0.05 可接受
Jarque-Bera (JB)	另一个常见的残差正态性检验	和 Omnibus 类似	$p > 0.05$ 理想
Skew (偏度)	残差对称性程度	正态分布偏度应接近 0	在 $[-1, 1]$ 之间通常可接受
Kurtosis (峰度)	残差的尖锐程度	正态值为 3	在 $[2, 4]$ 之间一般认为合理
Durbin-Watson	检验残差是否自相关 (自回归)	检查是否有序列相关性	约为 2 表示无自相关, 1.5 - 2.5 接受

9. 举例说明

```

Dep. Variable: FNE
R-squared: 0.503
Adj. R-squared: 0.502
coef(SA): 0.7092
P>|t|: 0.000
Durbin-Watson: 2.033
Skew: 0.233
Kurtosis: 4.164

```

模型可解释 FNE 的约 50.3% 变异，属于较强的解释力 ($R^2 = 0.503$)。SA 显著预测 FNE ($p < .001$)，其标准化回归系数为 0.709，表示 SA 每上升 1 单位，FNE 增加约 0.71 单位。残差近似正态 (偏度 ≈ 0.233 ，峰度 ≈ 4.16)，自相关可接受 (DW ≈ 2.03)。