R包piecewiseSEM的分段结构方程建模

常规的结构方程模型（SEM）局限性较大，如要求数据尽量服从多元正态性等，限制了SEM的使用。

分段结构方程模型（分段SEM，或称验证路径分析）通过引入灵活的数学框架扩展了SEM的适用范围。该数学框架允许合并各种模型结构、分布和假设，例如交互作用和非正态响应、随机效应和层次模型以及其它相关结构（包括系统发育、空间和时间）等。因此与常规的SEM相比，更多的变量间关系在分段SEM中更容易估计。

关于常规SEM和分段SEM的基本概念描述，可[参考前文](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484496&idx=1&sn=f0cc6f19bf5ea6861aeb38ba8195954d&chksm=97f5b448a0823d5eead0d8739d661a2e680e01e47c1432a345c23a4e54a7df463f64ceba61ba&token=1398584337&lang=zh_CN" \l "rd)。

本篇简介R包piecewiseSEM的分段SEM方法。

## 数据集

piecewiseSEM的内置数据集keeley，记录了加利福尼亚某些地区遭受火灾之后，火灾地区的位置以及植物（灌木林）多样性等信息。



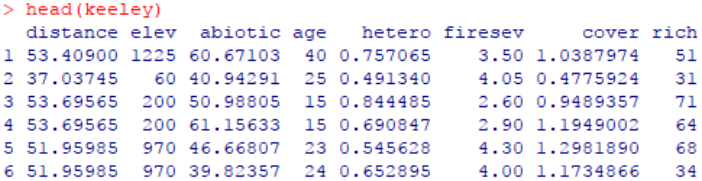
（来源：<https://pubs.er.usgs.gov/publication/1008266>）

library(piecewiseSEM)

#数据集，详情 ?keeley

data(keeley)

head(keeley)



distance，地区到海岸的距离；

elev，海拔高度；

abiotic，非生物条件；

age，灾前的林分年龄；

hetero，地区内异质性

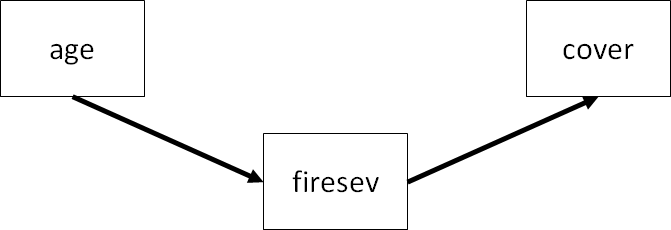
firesev，火灾严重性；

cover，灌木覆盖度；

rich，灌木物种丰富度。

研究人员希望了解火灾程度和林分年龄的关系，以及灾后灌木物种多样性的恢复模式：林分年龄是否导致更严重的火灾程度，火灾程度严重的地区灾后灌木丛恢复能力是否更低。

也就是构建如下模型：



如果使用分段SEM，则需分别构建两组方程：

firesev ~ age，

cover ~ firesev。

所谓“分段”，即每组关系都是独立估计的，最后合并以生成有关全局SEM的推论。

## piecewiseSEM的分段SEM

接下来，我们使用piecewiseSEM包中的分段SEM方法建模。

线性回归是大多数情形下直接考虑的方法，此时拟合分段SEM过程中即包含了各组独立的线性回归模型。

#分段 SEM，详情 ?psem

#这里，对于每个独立的响应方程，直接使用简单线性回归，确定响应关系

#其它情况，如果已知变量间的某种非线性关系，即可以使用非线性模型

keeley\_psem <- psem(

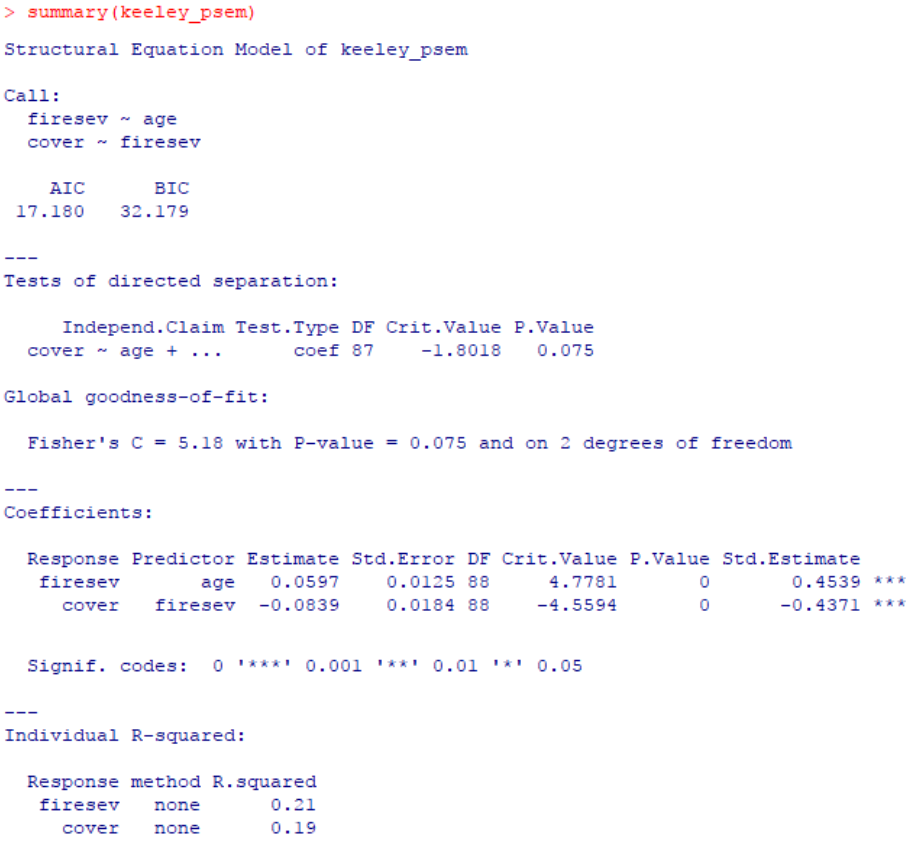
lm(firesev ~ age, data = keeley),

lm(cover ~ firesev, data = keeley),

data = keeley)

summary(keeley\_psem)

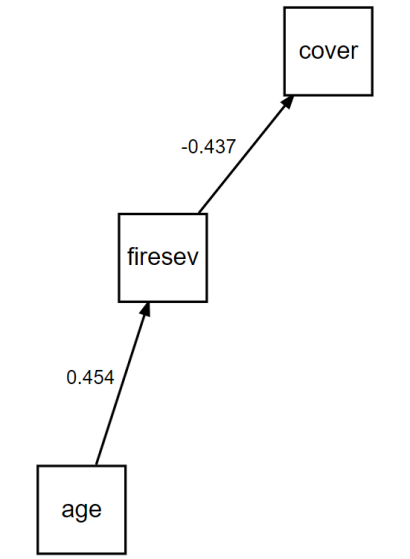
plot(keeley\_psem)



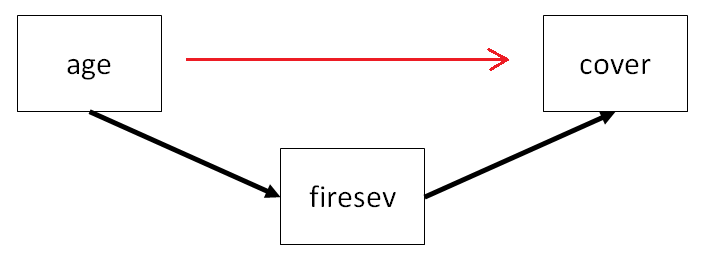
概要中包括了分段SEM的拟合度评估指标，以及表征变量因果关系的参数估计值及其显著性等信息。

从概要中我们可以获知，两组独立的线性回归模型显著，表明更老的灌木林以更严重的大火为特征，并与灾后植物覆盖率的低恢复率和丰富度有关。

路径图展示了模型中变量间的因果关系，图中的数值为标准化后的参数估计值（标准化的回归系数）。



那么另一问题，林分年龄和灾后植物恢复率之间是否存在直接的显著因果关系（二者之间是否也可以得到一个直接指向）呢？也就是如下模型：



根据上述概要所示Fisher’s C统计量的p值大于0.05，表明原模型中的结构是合理的，原模型中已经识别的所有有效路径，不推荐再添加更多路径。（[有向分离测试](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484496&idx=1&sn=f0cc6f19bf5ea6861aeb38ba8195954d&chksm=97f5b448a0823d5eead0d8739d661a2e680e01e47c1432a345c23a4e54a7df463f64ceba61ba&token=2137930916&lang=zh_CN#rd)评估）

另一种方法，通过[Akaike信息准则](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484496&idx=1&sn=f0cc6f19bf5ea6861aeb38ba8195954d&chksm=97f5b448a0823d5eead0d8739d661a2e680e01e47c1432a345c23a4e54a7df463f64ceba61ba&token=2137930916&lang=zh_CN#rd)（AIC）评估。由于AIC通常在比较多种模型时使用，因此我们不妨创建含age和cover关系的新模型，并比较它和原模型的AIC值水平。或者，通过全局的卡方检验比较新旧模型的差异是否显著。

#将 age 和 cover 的因果关系考虑在内，创建新模型

keeley\_psem2 <- psem(

lm(cover ~ firesev + age, data = keeley),

lm(firesev ~ age, data = keeley),

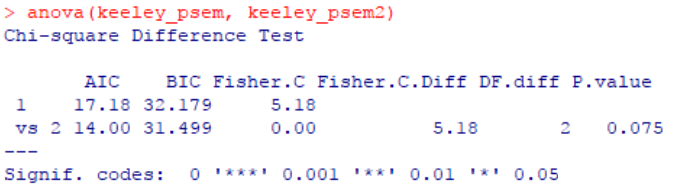
data = keeley

)

#比较新（含 age 和 cover 的关系）旧（不含 age 和 cover 的关系）模型

#全局模型的卡方检验

anova(keeley\_psem, keeley\_psem2)



尽管AIC、BIC值等显示新模型优于旧模型，但似乎二者数值差异并不很大，暗示新模型相对于旧模型，整体拟合度没有较大的提升。并且，新旧模型间全局的卡方检验p值不显著，表明两个模型没有显著差异。综合考虑，倾向于选择简约性的原模型。

此外，对于主要统计量，尽管它们均在summary()概要中有所展示，但也可使用特定函数单独查看。

#参数估计值（回归系数）

coefs(keeley\_psem, standardize = 'none')

coefs(keeley\_psem, standardize = 'scale') #可显示标准化后的

#定向分离测试

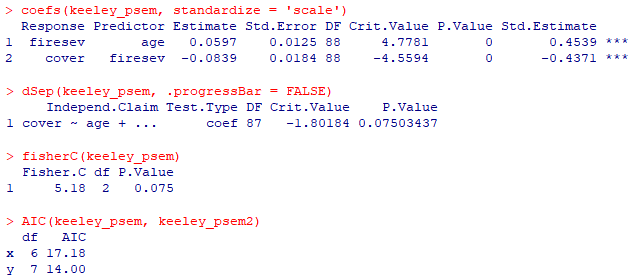
dSep(keeley\_psem, .progressBar = FALSE)

#Fisher’s C 统计量，p 值等同于上述定向分离测试

fisherC(keeley\_psem)

#AIC 值，用于比较多个模型

AIC(keeley\_psem, keeley\_psem2)



上述过程展示了在分段SEM中使用最简单的线性回归分步拟合模型。

实际应用中，变量间的关系往往比较复杂，有时可能还需引入较复杂的方法。

最后再顺便举一个小示例，在分段SEM中使用广义线性模型。

#一个带广义线性模型的 SEM 示例

#直接用随机数展示在分段 SEM 中使用到 glm 时的书写方式了

set.seed(1)

dat <- data.frame(

x = runif(100),

y1 = runif(100),

y2 = rpois(100, 1),

y3 = runif(100)

)

modelList <- psem(

lm(y1 ~ x, data = dat),

glm(y2 ~ x, family = 'poisson', data = dat),

lm(y3 ~ y1 + y2, data = dat),

data = dat

)

summary(modelList, conserve = TRUE)

## 参考资料

<https://jslefche.github.io/sem_book/local-estimation.html>