

线性混合模型在重复测量资料中的应用

作者：陶立元 赵一鸣 来源：临床流行病学和循证医学

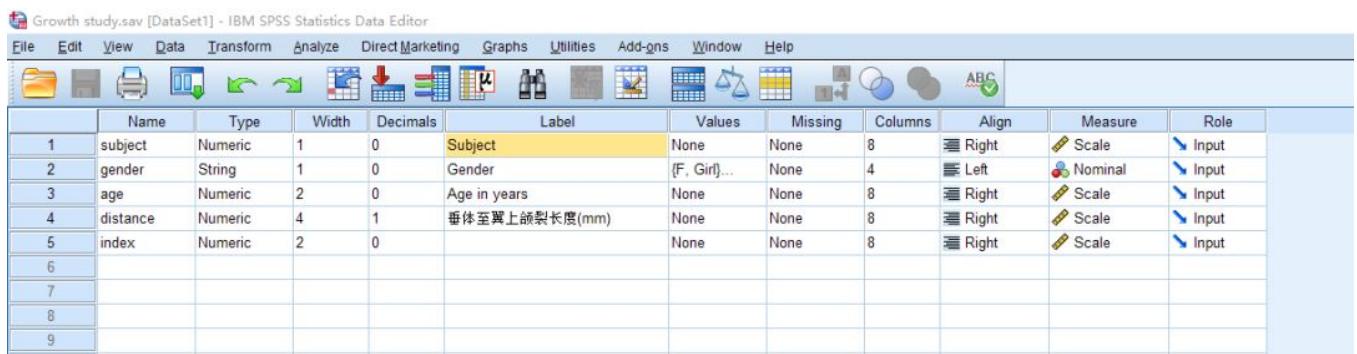
本文原地址：<https://www.iikx.com/news/statistics/1465.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

重复测量资料在我们的临床研究中经常遇到，如果数据符合独立正态和方差齐，我们可以采用重复测量的方差分析去解决他。同样我们还可以采用混合线性模型去进行分析，本次咱们就讨论一下混合线性模型在重复测量方差分析中的应用。

混合线性模型是20世纪80年代初针对统计资料的内部相关而发展起来的一种统计方法。一般意义上，变异的来源主要包括固定效应、随机效应和混合效应，所谓的固定效应是指分析的内容仅是目前的分组，不作外推的打算。比如ABC三组的比较，分析的目的仅仅是这三组间的差异，而不是超越这三组以外的差异。而随机效应是指分析的目的不仅仅局限在分组以内，比如北京、长沙和洛阳某指标的比较，研究者不仅想比较这三个城市，还想依次外推到一线、二线和三线的城市的情况。混合效应模型则是固定效应+随机效应模型，最典型的混合效应模型就是线性混合效应模型，除此之外还有广义线性混合效应模型。再说一下线性模型和广义线性模型，线性混合模型一般要求y符合或接近符合正态分布，多元线性回归、方差分析和协方差分析均是线性模型的经典应用;而广义线性混合模型则是线性模型的普及化，它是一种非正态因变量的回归模型。

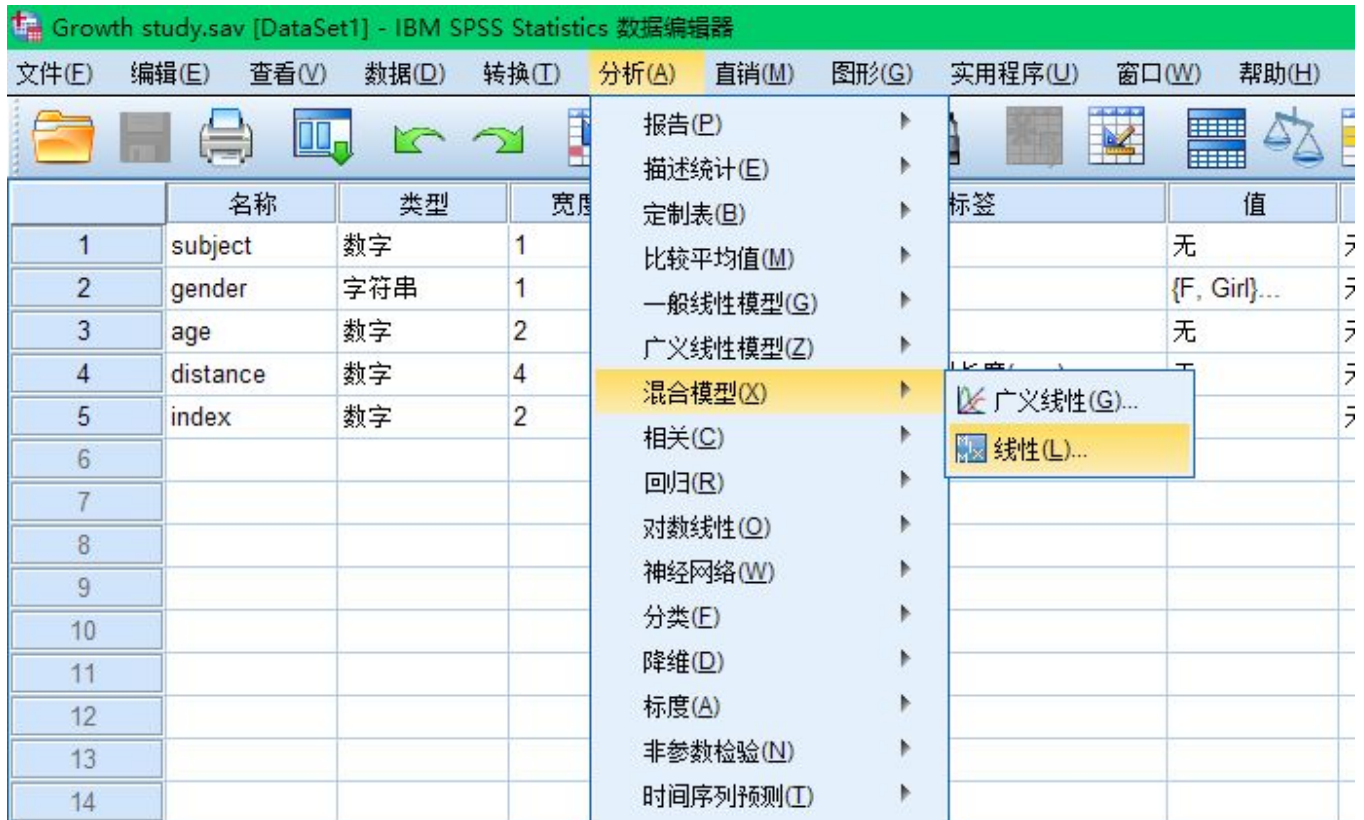
讲了上面的这些基本情况，咱们来看一个混合线性模型处理重复测量资料的例子。用SPSS自带的数据库Growth study.sav。该数据库为了研究青少年的牙齿发育情况与年龄和性别的关系，随机选取了27名儿童，通过随访观察，分别在其8岁、10岁、12岁和14岁时测量了某口腔指标，起名为distance。数据库的变量内容如下图：



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	subject	Numeric	1	0	Subject	None	None	8	Right	Scale	Input
2	gender	String	1	0	Gender	{F, G}...	None	4	Left	Nominal	Input
3	age	Numeric	2	0	Age in years	None	None	8	Right	Scale	Input
4	distance	Numeric	4	1	垂直至翼上颌裂长度(mm)	None	None	8	Right	Scale	Input
5	index	Numeric	2	0		None	None	8	Right	Scale	Input
6											
7											
8											
9											

这是一个按时间顺序的重复测量数据，又称纵向数据(longitudinal data)。其SPSS操作步骤如下：

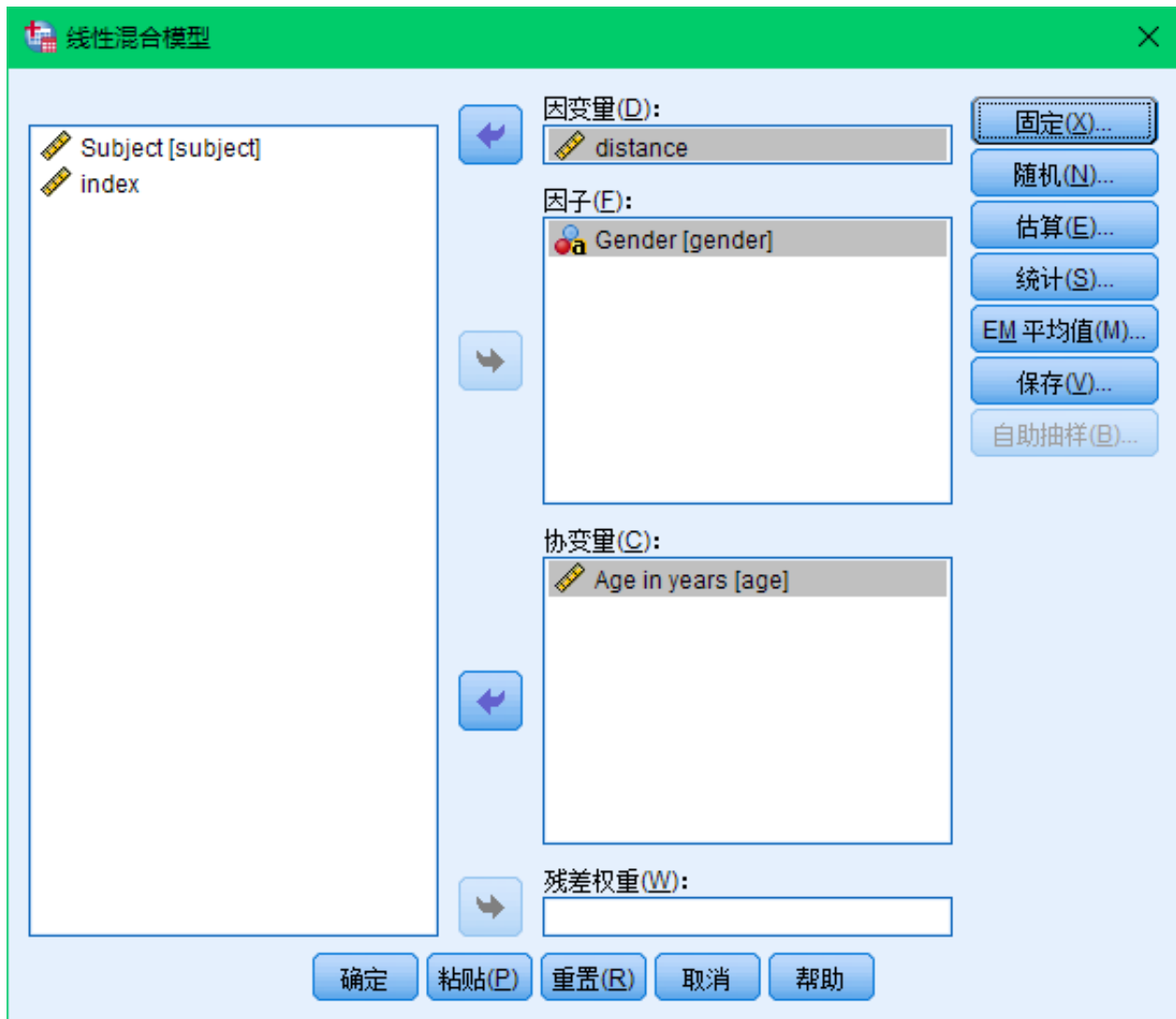
1、点击打开线性混合模型对话框



2、放入对象和重复测量变量，在重复协方差类型中选择“标度恒等(scaledidentity)”，点击“继续”。



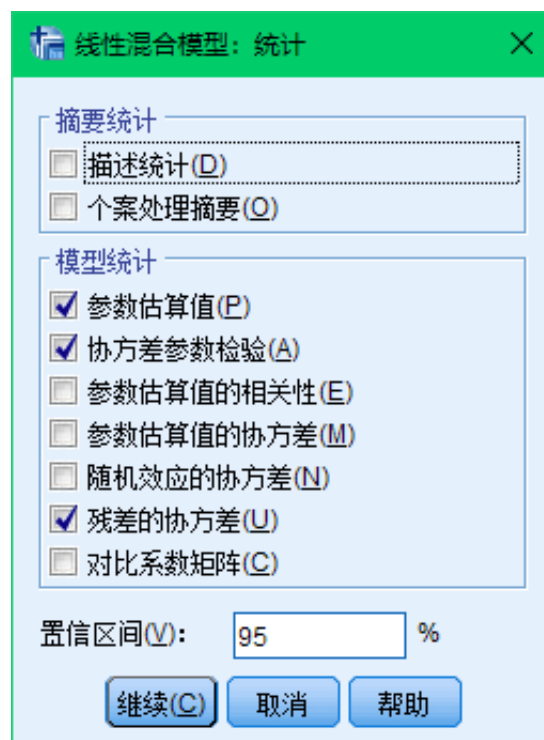
3、将distance选入因变量，将性别选入因子，age选入协变量



4、点击“固定”对话框，将性别和年龄的主效应放入模型，后点击“继续”



5、然后在统计中选出如下项目后，点击“继续”。再点击确定即可开始计算。



这时候我们可以在计算结果中看到如下结果，即固定效应性别和年龄均是有差异的，其中性别的 $p=0.01$ ，年龄的 $p<0.001$ 。

固定效应

III 类固定效应检验^a

源	分子自由度	分母自由度	F	显著性
截距	1	24.108	404.711	.000
gender	1	25.000	7.719	.010
age	1	26.000	92.284	.000

a. 因变量：distance。

固定效应估算^a

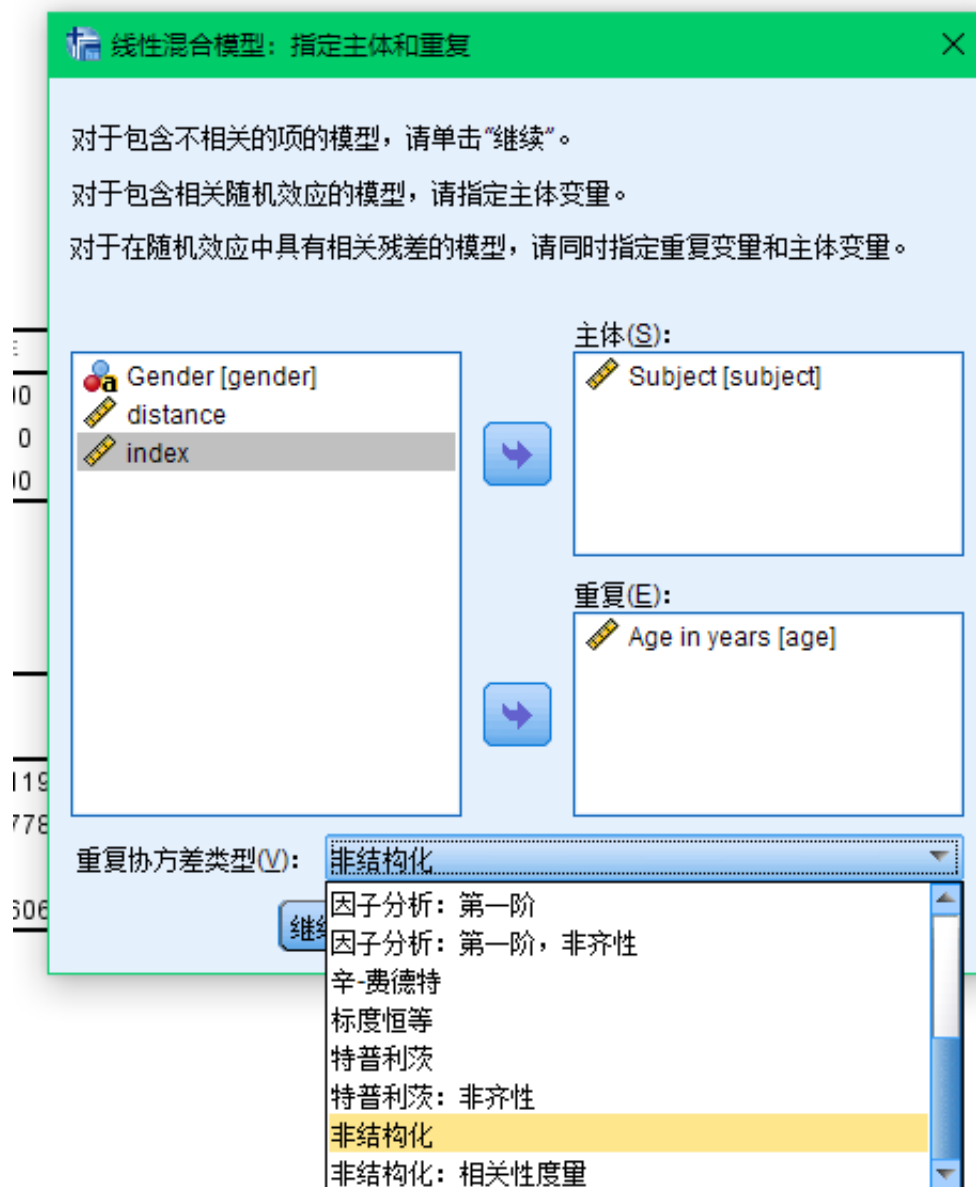
参数	估算	标准误差	自由度	t	显著性	95% 置信区间	
						下限	上限
截距	17.417595	.865718	28.785	20.119	.000	15.646429	19.188761
[gender=F]	-2.045173	.736141	25.000	-2.778	.010	-3.561284	-.529062
[gender=M]	0 ^b	0
age	.674651	.070229	26.000	9.606	.000	.530293	.819008

a. 因变量：distance。

b. 此参数冗余，因此设置为零。

至此该重复测量资料的线性混合模型便计算完成了。但是在计算过程中有一个关键问题，就是“重复协方差类型”我们选择为非结构化，如下图。通过下拉菜单我们发现它有很多个选择，不同的选择代表着对于相邻两次测量之间的相关性采用不同的估计方法。SPSS中相关性的估计有独立（即不相关）、等相关（即假定任意两时点的相关是相等的）、平稳相关（即间隔时间长度相等的测量间相关相同）、自相关（相关系数随着间隔次数的不同而不同）和非确定性相关（即不做任何限制，任意两点之间的相关都可能不同）。

具体协方差的结构的结构的选择，可以参照SPSS的帮助文档，有详细的说明。同时结合自己数据的特征，选择合适的协方差结构，较为常用的协方差结构是等相关和一阶自相关。



参考书目：张文彤、董伟主编，《SPSS统计分析高级教程》。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发